Students' attitudes and intentions towards studying science: the effects of under-confidence and over-confidence

Richard Sheldrake

UCL

PhD

Declaration

I, Richard Sheldrake, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Richard Sheldrake

Word count (excluding front matter, references, and appendices): 81,844

Abstract

Understanding students' intentions to study science at upper-secondary school continues to be a central concern for science education. Prior research has associated students' confidence with their intentions to study science, although under-confidence and over-confidence (lower or higher confidence than expected given someone's attainment) has not been considered in detail. Under-confident students may not select subjects that they might otherwise succeed in and enjoy, which may be a fundamental barrier. Accordingly, this study explored whether under-confident, accurately-evaluating, and over-confident students expressed different attitudes towards their science education, and explored how under-confidence and over-confidence might influence students' science intentions. Existing nationally-representative data and newly-collected data from secondary school students in England were considered in order to provide complementary insights and to enhance the plausibility of the findings. Multiple analytical approaches were applied to consider under-confidence and over-confidence, including calculating various indicators of accuracy/bias and applying various approaches to grouping students. The results highlighted that under-confidence and over-confidence may be problematic, not simply through associating with lower or higher attitudes, but also through students considering their choices in different ways.

Table of contents

List of tables

List of figures

Section 1: Introduction

Understanding students' intentions to study science at upper-secondary school continues to be a central concern for science educators in England (Royal Society, 2014). The numbers of students studying science, and related subjects, at upper-secondary school and at university in England have historically varied, and have often been lower than other subjects and imbalanced across boys, girls, and students with different backgrounds (Royal Society, 2006, 2008a; Smith, 2011). More students studying science have been desired as a means to foster greater quantitative skills (British Academy, 2015; Leitch, 2006; OECD, 2015b), to meet an expected demand for increased numbers of scientists and science-related professionals (Bosworth, Lyonette, Wilson, Bayliss, & Fathers, 2013; Roberts, 2002; Wilson, Beaven, May-Gillings, Hay, & Stevens, 2014), and to address under-representation and promote equity (CASE, 2014; Institute of Physics, 2014; WISE, 2014).

Secondary school students in England have generally considered science to be fairly interesting and relevant for careers, although relatively few students have liked science better than other subjects or explicitly aspired to be scientists (DeWitt, Archer, & Osborne, 2014; Jenkins & Nelson, 2005). Concurrently, girls, those from families with low incomes, and those from Black, Bangladeshi, and Pakistani backgrounds have been less likely to study upper-secondary science subjects, while boys, those from families with high incomes, and those from Chinese and Indian backgrounds have been more likely, although it remains difficult to isolate particular causes (Department for Education, 2011; Institute of Physics, 2014; Royal Society, 2008b).

Across many research studies in England and other countries, students' attitudes towards science, such as their interest in science and perceived utility of science, and their motivational beliefs such as their confidence in their own abilities, together with their own attainment, have closely associated with their intentions and choices (Bøe & Henriksen, 2015; Regan & DeWitt, 2015). Indeed, students' intentions to study science have been predicted more by their own attitudes and beliefs than by their background and characteristics such as their gender (DeWitt, Archer, &

Osborne, 2014; Mujtaba & Reiss, 2014). Essentially, and as proposed by motivational theories (Eccles, 2009), various aspects of students' background and context, such as their parents' beliefs (DeWitt, et al., 2011) and classroom experiences (Wang, 2012), may influence their attitudes about science, which may then primarily influence their intentions.

Differences in students' attitudes and beliefs may then help explain both progression and under-representation in science. For example, attainment in itself does not appear to be relevant to the low representation of girls, but girls often report lower science attitudes and confidence in their abilities (Mujtaba & Reiss, 2013; OECD, 2015a; Wang & Degol, 2013). Intuitively, and as shown by some interventions, promoting positive attitudes towards science, such as higher perceptions of the utility of science, may then help increase the number of students studying science (Harackiewicz, Rozek, Hulleman, & Hyde, 2012; Rozek, Hyde, Svoboda, Hulleman, & Harackiewicz, 2015).

Students' confidence, however, appears to require closer consideration. Students' confidence does not necessarily correspond to their actual attainment: reviews have found only modest associations between various indicators of each (Freund & Kasten, 2012; Hansford & Hattie, 1982; Ma & Kishor, 1997; Mabe & West, 1982; Zell & Krizan, 2014). Further studies have highlighted that students can be under-confident, with lower confidence than would be expected given their attainment, while others can be over-confident, with higher confidence than would be expected given their attainment (Bouffard & Narciss, 2011). While higher confidence may be motivationally beneficial (Bandura, 1997), aiming to increase the number of students studying science through universally increasing confidence may reduce under-confidence for some but further increase over-confidence for others, and it is unclear whether this would be helpful.

Students' confidence, expressed in various ways, has nevertheless closely associated with their studying intentions (Bong, 2001b; Mujtaba & Reiss, 2014). Promoting equity entails understanding potential barriers before they can be addressed; students can then, ideally, make informed and unconstrained choices about whether to study science or not. Under-confidence may potentially limit students' progression or identification with science, but this may be an avoidable barrier; conversely, over-confidence may entail subsequent problems if students lack sufficient attainment for their future plans.

Under-confidence and over-confidence have not been extensively explored within science education, however. For example, studies have found that higher confidence has associated with higher science intentions when controlling for attainment (Mujtaba & Reiss, 2014). While this may lead to the inference that higher confidence is beneficial, even if someone is over-confident, such results do not isolate under-confidence or over-confidence from low or high confidence in itself. For some students, confidence and attainment may be sufficiently similar to entail accuracy; for other students, differences may be large enough to entail under-confidence or over-confidence. Similarly, prior research has not considered whether any revealed patterns of predictive associations with intentions occur regardless of whether someone is under-confident, accurate, or over-confident in their beliefs. Instead, other methods are required to first identify students with different confidence biases and then to explore their expressed beliefs and their patterns of associations within predictive models.

Accordingly, this thesis identifies under-confident, accurately-evaluating, and over-confident students, via various approaches, in order to explore how these cases might be detrimental or beneficial within science education. For example, under-confident students might report lower attitudes towards science, including for factors that predict intentions to study science further. Additionally, the thesis considers whether students' science intentions are predicted in different ways, depending on whether students are under-confident, accurately-evaluating, or over-confident. Any differences would provide greater understanding into how students' choices are made, and further insights into the potential impact of under-confidence or over-confidence.

Section 1.1: About this thesis

This thesis developed from, and extends, earlier research associated with mathematics progression in England (Sheldrake, Mujtaba, & Reiss, 2014,

2015) undertaken to complement a wider national research project covering science and mathematics (Reiss, et al., 2011; Mujtaba & Reiss, 2014). Specifically, this earlier research revealed that, for Year 8 students, under-confident students reported lower mathematics attitudes than accurate and over-confident students, including for their mathematics interest and perceived utility of mathematics; at Year 10, fewer differences were observed, but over-confident students reported the lowest intentions to study mathematics further (Sheldrake, Mujtaba, & Reiss, 2014). Indicators of confidence accuracy/bias were also directly predictive of students' mathematics intentions, and there were some indications that intentions could be predicted differently for various groups of students although this was not statistically confirmed (Sheldrake, Mujtaba, & Reiss, 2015).

The situation remained unclear for science, however. The majority of research into students' confidence accuracy/bias has covered various academic subjects with less focus on science, has been undertaken outside of England, and students' studying intentions have seldom been concurrently considered (Cheema & Skultety, 2016; Chen, 2003; Chen & Zimmerman, 2007; Gonida & Leondari, 2011; Rinne & Mazzocco, 2014). Additionally, emerging methods to consider confidence accuracy/bias through clustering students appeared not to have been undertaken in England (Bouffard, Vezeau, Roy, & Lengelé, 2011; Rytkönen, Aunola, & Nurmi, 2007; Sáinz & Upadyaya, 2016; Seidel, 2006).

The research underlying this thesis accordingly aimed to apply multiple complementary approaches in order to explore the impact of confidence accuracy/bias on students' intentions, attitudes, and beliefs, specifically considering science, students in England, and including approaches such as clustering students. Science was considered holistically as in the National Curriculum for England, encompassing the natural sciences of biology, chemistry, and physics (Department for Education, 2013, 2014). For contextualisation, wider literature was also considered across STEM (Science, Technology, Engineering, and Mathematics) subjects, and other areas, when this potentially added insight and/or when little research had directly considered science or confidence accuracy/biases.

Existing data from the Programme for International Student Assessment (PISA) 2006 was analysed, which surveyed a nationally-representative sample of Year 11 students in England and which considered a broad array of students' attitudes and beliefs in science, including their studying intentions (OECD, 2009a). However, the PISA design ensured that the students' confidence accuracy/bias could only be broadly approximated. Accordingly, new data were collected (2014/2015) and analysed in order to consider students' confidence accuracy/bias in more detail for students in Years 9, 10, and 11, using similar methods to prior research undertaken outside of England (Chen, 2003; Chen & Zimmerman, 2007). The new survey considered an array of attitudes and motivational beliefs, but could only reach a relatively small number of students given limited resources. Similarities and differences across the two sets of results were then considered; specifically, any similarities in results across the different samples and different analytical approaches could then enhance the overall plausibility of any emerging findings.

Initial analysis, not presented within this thesis for brevity, provided wider contextualisation and explored potential antecedents of confidence biases, which had surprisingly received little explicit attention within earlier research. Findings included that, for example, lower subject-comparisons (science thought to be harder than any other subject) predicted lower self-concept beliefs (confidence conceptualised as subjective subject-level beliefs of personal ability) when students were under-confident but not when they were over-confident, for the Year 9, 10, and 11 students surveyed in 2014/2015 (Sheldrake, 2016a) and for Year 9 students in the Trends in International Mathematics and Science Study (TIMSS) 2011 (Sheldrake, 2016b). Analysis then focused on the main area of students' science intentions. Initial findings highlighted that, for the surveyed Year 9, 10, and 11 students, science intentions were predicted by different factors in different ways, confirmed with statistical significance, depending on whether they tended to under-confidence, accuracy, or over-confidence (Sheldrake, 2016c).

This thesis provides further insight, and reports on multiple analytical approaches and both the PISA 2006 survey and the 2014/2015 survey, including considering confidence accuracy/bias by clustering students. Similarities and differences across the approaches and surveys were considered in order to enhance the plausibility of any findings, to help address methodological limitations in one set of data with strengths in the other, and to ultimately provide new insights.

Throughout the thesis, the various terms aimed to be intuitive and inclusive, where possible. For example, 'attainment' has generally been used to refer to students' performance, achievement, and/or similar terms. Similarly, 'confidence' has been used to broadly encompass terms such as students' self-concept (subjective beliefs of past/current ability) and self-efficacy (evaluative beliefs of future capacity/capability), although the particular terms have also been used when relevant.

Section 1.2: Contents and structure

Prior research into students' intentions and choices is briefly reviewed in **Section 2**, together with the theoretical expectancy-value model of students' motivated behavioural choices. This broadly highlighted the influence of students' confidence on their intentions and choices, although the impact of under-confidence or over-confidence remained unclear. Students' confidence is then considered in more detail in **Section 3**, including different conceptualisations and supporting theory; specific studies into under-confidence and over-confidence are also detailed.

The research aims and the overall design of the research are then described in **Section 4**, including the three underlying research questions. The research considered nationally-representative data from PISA 2006, and new data collected in 2014/2015, and the particular methods are described in **Section 5** for both surveys. Analytical approaches are then elaborated in **Section 6**, such as how missing responses were handled and how analytical modelling was undertaken.

The results for the three research questions are respectively described and discussed in **Section 7**, **Section 8**, and **Section 9**. For each research question, the section covers the results from both surveys, discusses similarities and differences, and contextualises the findings against prior research. **Section 10** then provides a general discussion, focusing on providing an overall summary of new insights, and also considering limitations and implications to future research. References and appendices then follow.

Section 2: Students' intentions and choices

The following sections briefly review influences on students' intentions and choices in **Section 2.1**, and further aspects related to students' attitudes and beliefs in **Section 2.2**. Specifically, low science intentions have been considered to follow from low attitudes towards science, generally declining as students grow older (**Section 2.2.1**); differences in intentions may follow from different students holding different attitudes or beliefs (**Section 2.2.2**); and differences in intentions may follow from different students considering their intentions in different ways (**Section 2.2.3**). A plausible theoretical framework for students' intentions, the expectancy-value model of motivated behavioural choices, which highlights the importance of students' confidence, is then considered in **Section 2.3**.

Section 2.1: Influences on students' intentions and choices

In England, science and mathematics are not compulsory subjects in upper-secondary education (EACEA, 2011; Eurydice, 2016). Studying science subjects at upper-secondary school is generally necessary to study science courses at university, however, and studying science at university is generally necessary for a science career (Royal Society, 2006). In England, relatively early experiences or choices at secondary school may then become even more important in facilitating or precluding future careers in science. Intuitively, it then remains important to gain a wider understanding of secondary school students' intentions and choices to study science further.

Students' aspirations and intentions to continue with science, reported during secondary school, have been shown to be important indicators of their future educational progression. For example, large-scale nationally-representative longitudinal studies in the United States of America, such as the National Educational Longitudinal Study starting in 1988 and the Education Longitudinal Study starting in 2002, have revealed that students' views and intentions to study science reported during secondary school indeed predicted whether they subsequently gained science degrees (Maltese & Tai, 2011; Morgan, Gelbgiser, & Weeden, 2013; Tai, Qi Liu, Maltese, & Fan, 2006; Wang, 2013). Similarly, in England, national longitudinal studies have highlighted that science aspirations reported during secondary school have associated with subsequent employment in science (Schoon, 2001).

Various factors have been found to associate with students' intentions or choices, although it remains difficult to determine their relative magnitudes or effects. For example, various studies have highlighted associations between intentions/choices and students' background or characteristics such as gender or ethnicity, or students' educational context such as characteristics of their school, although often without considering students' own views (Bennett, Lubben, & Hampden-Thompson, 2013; Crawford, 2014; Gill & Bell, 2013; Homer, Ryder, & Banner, 2014). However, differences in the number of students studying Advanced Level General Certificate of Education (A-Level) science subjects across students' backgrounds and types of schools have been shown to the small, after controlling for the students' prior attainment and gender (Gill & Bell, 2013; Homer, Ryder, & Banner, 2014). Further research has revealed that students' intentions to study science have been predicted more by their own attitudes and beliefs than by their background, gender, or attainment, although the various effects have slightly varied across studies and sets of considered predictors (DeWitt, Archer, & Osborne, 2014; Mujtaba & Reiss, 2014).

The influence of factors may still be complex and difficult to disentangle. For example, students' attainment in secondary school, such as in General Certificate of Secondary Education (GCSE) or equivalent examinations, has often been highlighted to be relevant to selecting A-Level subjects and attaining high grades in them (Department for Education, 2011, 2012; Homer, Ryder, & Banner, 2014). However, schools may require particular levels of attainment in order to permit students to study upper-secondary sciences (Department for Children, Schools and Families, 2009). It can be difficult to determine whether some students are unable to continue to study particular subjects at A-Level at their particular school, or whether they explicitly decide not to study them. Additionally, students' intentions have associated with their predicted or expected grades, which may differ from their current or subsequent attainment (Brown, Brown, & Bibby, 2008).

Prior research has frequently highlighted that intentions and/or choices have associated with students' attainment and also their beliefs in their own abilities (their confidence), together with their intrinsic interest in a subject and their extrinsic career concerns such as gaining a specific job or gaining a well-paid job (Bates, Pollard, Usher, & Oakley, 2009; Blenkinsop, McCrone, Wade, & Morris, 2006; Brown, Brown, & Bibby, 2008; Cann, 2009; Department for Education, 2012; McCrone, Morris, & Walker, 2005; Regan & DeWitt, 2015; Tripney, et al., 2010). More specifically, commonly reported reasons for studying science and related subjects at upper-secondary school in England have included the perceived usefulness of the subject, perceived ability in the subject, enjoyment, the complementary nature of some subjects, and interest (Tripney, et al., 2010). For example, and in more detail, students' science aspirations at the end of primary school (Year 6) and in the middle of secondary school (Year 9) were more strongly predicted by their relevant attitudes and beliefs (students' attitudes to science, parents' attitudes to science, and confidence expressed as self-concept beliefs in science) than by their background characteristics (gender, ethnicity, parental education, books at home, and parents working within science) (DeWitt & Archer, 2015). Similarly, for Year 10 students in England, students' intentions to continue to study physics were most strongly predicted (in descending magnitudes) by their perceived utility of physics, interest in physics, home support for achievement in physics, emotional responses to physics lessons, perceptions of physics lessons, physics confidence expressed as self-concept beliefs, and their perceived advice/pressure to study physics (Mujtaba & Reiss, 2014). For Year 12 and 13 students, A-Level choices across various subjects were retrospectively reported to follow from their interest and enjoyment, perceived utility for careers, and their own perceived ability; students cited utility more for science and mathematics related subjects and cited interest more for humanities subjects (Vidal Rodeiro, 2007).

Considering subject choices for university in general, interest in a subject, then career concerns (gaining a specific job and gaining a well-paid job), and then being good at a subject were considered to be most relevant by students at the end of secondary school in England; the opinions of family members and the ease of getting onto a course were considered to be far less important (Bates, Pollard, Usher, & Oakley, 2009). University students studying science and related subjects in England have similarly expressed that they originally selected a science subject because of interest, enjoyment, or ability, rather than primarily for career-related reasons; those who cited career-related reasons highlighted that science would also keep their options open (Mellors-Bourne, Connor, & Jackson, 2011). Science graduates in the United States mostly reported that their own interest was the strongest influence that maintained their persistence in science and related subjects during middle school, high school, and college (Maltese, Melki, & Wiebke, 2014).

Similarly, university students studying science and related subjects across England, Denmark, and Norway most frequently mentioned that they selected their courses due to interest and enjoyment (mentioned in 68% of their written responses), then utility (17%), expectancies of success (16%), and then the personal value of the course area to their own identities (9%) (Jensen & Henriksen, 2015). Other influences, such as school experiences including teachers, family, popular science and leisure activities, and outreach activities, most frequently linked with students' expressed interest and enjoyment (Jensen & Henriksen, 2015). When considering only the potential influences of other people, university students of science in Norway most frequently reported that their parents were influential on their choices to study science, then friends, other acquaintances, teachers, siblings, and (to a far less frequent extent) public figures in the media (Sjaastad, 2012). Fathers were mentioned more than mothers or both parents together (Sjaastad, 2012).

Students' narratives about their university choices in Denmark similarly highlighted that their choices were broadly based on their interests, but balanced by other factors including the implications of being a student on a particular course and the likely future employment situation (Holmegaard, Ulriksen, & Madsen, 2014). Aspects of studying subjects or perceptions of wider fields had different implications for different students. For example, some who decided to study science were attracted to the rigorous methods and clarity in science, while others who did not study science were disenchanted by perceiving science as rigid and superficial (Holmegaard, Madsen, & Ulriksen, 2014).

Considered in general terms, students' conceptions of themselves or their identities, and students' conceptions of science and/or scientists, have appeared to be relevant to students' intentions and choices (Andersen, Krogh, & Lykkegaard, 2014; Archer, et al., 2010; Bøe & Henriksen, 2015; Taconis & Kessels, 2009). For example, for secondary school students in England, differences between perceptions of science and students' perceptions of their own identities may entail that science is not considered as a feasible option for further study or careers, perhaps especially for girls (Archer, et al., 2013) and for those from particular backgrounds and ethnicities (Archer, DeWitt, & Osborne, 2015). Students have often associated scientists with generally being male and white, although often 'scientists' have only been interpreted to mean those working in the natural or physical sciences (Archer, DeWitt, & Osborne, 2015; Wong, 2015). Students have nevertheless recognised people with various backgrounds/ethnicities working in other science fields, such as medicine, which may ensure that aspiring to those fields can be more relatable (Wong, 2015). For secondary school students in the United States, the relations between aspects of students' identities and science, including similarities and differences, could associate with increasing or decreasing engagement, in various ways (Calabrese Barton, et al., 2013; Carlone, Johnson, & Scott, 2015; Tan, Calabrese Barton, Kang, & O'Neill, 2013). While elements of identity and their implications can be clearly highlighted through these case studies, it remains harder to determine their impact in comparison to other factors.

The influence of students' confidence on their intentions and choices has also been highlighted in various ways. In England, for example, secondary school students' confidence in science, measured as self-concept beliefs, has predicted their intentions to study science further, together with other factors (DeWitt & Archer, 2015; Mujtaba & Reiss, 2014). Students' self-concept beliefs have also associated with selecting science and mathematics courses at university in Germany (Parker, et al., 2012) and in Australia (Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014). In various countries, students' confidence, measured as self-efficacy beliefs during secondary schooling, has also been found to influence their ideas of potential careers (Bandura, Barbaranelli, Caprara, & Pastorelli, 2001), and to directly predict their intentions to study courses (Bong, 2001b) and to enter university (Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014). Similarly, during university studies, self-efficacy has also predicted students' intentions to complete their courses in science and related subjects (Byars-Winston, Estrada, Howard, Davis, & Zalapa, 2010). Women in mathematical, scientific, and technological careers in the United States have also highlighted that their strong self-efficacy enabled resiliency to overcome potential obstacles and gave a desire to succeed (Zeldin & Pajares, 2000). An increased understanding of students' confidence, and potential biases towards under-confidence or over-confidence, may then help provide wider insights into students' science intentions.

In summary, students' perceived utility, interest and enjoyment, and their perceived ability in science (their confidence) have closely associated with their science intentions and choices (Jensen & Henriksen, 2015; Mellors-Bourne, Connor, & Jackson, 2011; Mujtaba & Reiss, 2014; Vidal Rodeiro, 2007). It remains unclear whether any one factor is the most influential, however. Studies highlighting that students cited interest as the primary reason for their choices (Bates, Pollard, Usher, & Oakley, 2009; Mellors-Bourne, Connor, & Jackson, 2011) may contrast with studies that highlighted the importance of the perceived utility of science (Mujtaba & Reiss, 2014; Vidal Rodeiro, 2007). It remains possible that methodological aspects may be relevant, such as potential differences arising from students rating their attitudes compared to researchers categorising students' expressions, or whether prospective or retrospective attitudes are considered; it also remains possible that different attitudes are more or less relevant at different times. Predictive models aim to isolate the independent contribution and relative magnitude of different factors, but any effects may ultimately operate in complex ways.

Section 2.2: Students' attitudes towards science

Given the importance of students' attitudes such as their interest, various studies have explored how they have developed or changed throughout their education (**Section 2.2.1**), and/or whether different students hold different beliefs (**Section 2.2.2**). It is also possible that different students consider different attitudes as being more or less relevant to their decisions of what to study (**Section 2.2.3**). Any or all of these areas may help explain why different students decide to study science or not.

Section 2.2.1: Differences in attitudes across secondary school

Students in primary school in England have generally enjoyed science but have not necessarily seen themselves as becoming scientists (Archer, et al., 2010; Silver & Rushton, 2008; Turner & Ireson, 2010). For such students, possessing a natural interest in science has been considered important in order to be good at science, and the students could perceive their peers as being 'science people' or not (Archer, et al., 2010). Primary school students also have perceived 'science in school' as less exciting and different to their ideas of 'real science' (Archer, et al., 2010; Zhai, Jocz, & Tan, 2014). While relatively few secondary school students in England have aspired to be scientists or liked science better than other subjects, science has been generally considered to be interesting, relevant for careers and gaining wider knowledge, and important for school and wider life (DeWitt, Archer, & Osborne, 2014; Jenkins & Nelson, 2005). However, science at school was considered to be hard and difficult, while being talented, invested, and interested in science and science careers were considered to be closely linked (Archer, DeWitt, & Osborne, 2015).

Declining attitudes towards science as students grow older have often been considered a fundamental cause of the low numbers of students studying science (Osborne, Simon, & Collins, 2003; Royal Society, 2008b). However, secondary school students have not necessarily held negative attitudes towards science (DeWitt, Archer, & Osborne, 2014; Jenkins & Nelson, 2005). Additionally, for a large survey of science graduates in the

United States, while over half reported that their interest in science developed before or during primary/elementary school, around a third nevertheless reported that their interest developed during secondary school (Maltese, Melki, & Wiebke, 2014). Interviews with graduate students and scientists in the United States revealed similar results (Maltese & Tai, 2010). Interest in science developing during secondary school does not easily fit with assumptions of declining interest over the duration of secondary school, which also highlights the problem of attempting to generalise across all students. Conversely, the problem of generalisation may be again highlighted since positive attitudes may not necessarily entail aspirations towards science for all students (Archer, et al., 2010).

Various aspects may be relevant to changing attitudes. Studies have considered whether changes in attitudes have followed from specific changes in context and environment, such as from primary school to secondary school. For example, small decreases in competence beliefs and attitudes have been observed following this transition in Germany (Arens, Seeshing Yeung, Craven, & Watermann, 2013), while different students reported increases, decreases, or their beliefs remained stable across this transition in the United States (Harter, Rumbaugh Whitesell, & Kowalski, 1992). The nature of teaching or the objects of inquiry within science education, potentially more ‘everyday’ at primary school and more abstracted at secondary school, may also be relevant to differences in expressed interest (Anderhag, et al., 2016; Tröbst, Kleickmann, Lange-Schubert, Rothkopf, & Möller, 2016). Students at different ages may even consider or form their attitudes in different ways. For example, a small study with primary school and secondary school students in Germany highlighted that younger students associated interest more with emotional experiences, while older students found discovering new knowledge and autonomy to be more important to their interest (Frenzel, Pekrun, Dicke, & Goetz, 2012).

Further studies have considered students for longer periods of time, and have highlighted that students’ attitudes in many academic subjects have often declined after primary school and across secondary school, although the changes have often involved largely-positive views declining to slightly-positive or neutral views (Archambault, Eccles, & Vida, 2010;

Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Nagy, et al., 2010). For example, in the United States, secondary school students' attitudes towards science declined from Grades 6 to 10, becoming relatively neutral at Grade 10 (Simpson & Oliver, 1990). Again in the United States, students' mathematics interest was seen to decline from primary/elementary school and across secondary school and upper-secondary school (Grades 1 to 12) to become relatively neutral, while the perceived importance of mathematics declined from Grades 1 to 9 but then increased from Grades 10 to 12 (Fredricks & Eccles, 2002).

Similarly, and more specifically, in Canada from Grades 5 to 11, students' interest in science and technology declined over time but remained overall positive; similar declines were observed for French (as a first language), mathematics, and physical education (Potvin & Hasni, 2014). On average, students' views increased over time for their perceived importance of outside-school science and technology, utility of outside-school science and technology for society, perceived difficulty of science and technology, and their attraction to science and technology studies and careers (Potvin & Hasni, 2014). Nevertheless, their confidence expressed as self-concept beliefs and their interest for in-school science and technology decreased (Potvin & Hasni, 2014). Compared to other subjects, science and technology increased the most in perceived importance and also in perceived difficulty, with the exception of being similar in perceived difficulty to mathematics (Potvin & Hasni, 2014).

Specifically in England, the number of students reporting that science was one of their favourite lessons decreased from Year 7 to Year 9 (Bennett & Hogarth, 2009). Students' interest and attitudes towards science subjects generally declined from Year 7 to Year 9 but then had less change, or even increased, to Year 11 (Bennett & Hogarth, 2009). Again in England, science aspirations and perceived parental attitudes to science increased from Year 6 (the end of primary school) to Year 8 (after the start of secondary school), while perceived peer attitudes to science and participation in science activities outside school decreased, and the students' self-concept beliefs in science and attitudes towards science remained similar (DeWitt, Archer, & Osborne, 2014). It was nevertheless highlighted that the increase in aspirations followed from the increase in students intending to study more science in the future and to have a job that used science, rather than intending to specifically become a scientist (DeWitt, Archer, & Osborne, 2014).

Any changes may vary across students, however. For example, in Germany, students' interest in mathematics decreased over their secondary school education, although varied by course: those on advanced mathematics courses reported high and stable interest, while those on basic courses reported lower and declining interest over time (Köller, Baumert, & Schnabel, 2001). In the United States, from Grades 1 to 12, distinct clusters of students with different trajectories of reading self-concept and utility beliefs could be identified, some declining slowly but remaining high or moderate, some declining to low magnitudes, and some declining then increasing again (Archambault, Eccles, & Vida, 2010). Again in the United States, for students followed from Grade 4 to university/college, clusters of students' trajectories of change in their mathematics self-concept, interest, and perceived utility (considered as importance and usefulness) could be identified, and those with generally high beliefs (which slowly decreased over time) were most likely to select a mathematically-intensive course at university (Musu-Gillette, Wigfield, Harring, & Eccles, 2015). Other clusters showed beliefs that decreased faster over time, were moderate and remained stable, or were moderate and slowly decreased (Musu-Gillette, Wigfield, Harring, & Eccles, 2015).

In summary, 'declining attitudes' across secondary education may not necessarily entail negative attitudes, and students' attitudes to science and attitudes to other subjects may similarly change over time (Potvin & Hasni, 2014). However, distinct clusters of students with different magnitudes or trajectories of attitudes may be observed (Musu-Gillette, Wigfield, Harring, & Eccles, 2015), although it remains unclear why or what wider factors might predict membership of such clusters. Even so, considering various clusters of students appears to be a promising method, which could potentially be as informative as attempting to disentangle influences or attitudes associated with students' background or other characteristics.

Section 2.2.2: Differences in attitudes across students

Various studies have explored whether different students hold different attitudes towards science and other subjects, although often focusing on explicit groups of students such as boys and girls. For example, the low number of girls studying subjects related to mathematics and the physical sciences, and their associated low attitudes, has frequently been highlighted (Blickenstaff, 2005; Institute of Physics, 2013; Murphy & Whitelegg, 2006; Royal Society, 2014), although ability in itself does not appear to be relevant to the low representation of girls (Wang & Degol, 2013).

Across students' various expressions of confidence and beliefs about their personal abilities, boys have generally reported higher beliefs than girls even though girls often attain equally or even higher than boys (Fredricks & Eccles, 2002; Marsh & Yeung, 1998; Marsh, 1989; Rhodes, Roffman, Reddy, & Fredriksen, 2004; Skaalvik & Skaalvik, 2004; Watt, 2006; Wigfield, Eccles, Mac Iver, Reuman, & Midgley, 1991; Young & Mroczek, 2003). Research reviews have nevertheless found that gender differences in attitudes to mathematics, on average, had small magnitudes and were similar to differences in mathematics performance; when differences occurred, girls held less positive attitudes than boys (Hyde, Fennema, Ryan, Frost, & Hopp, 1990). Similarly, reviews of early international PISA and TIMSS studies have highlighted that boys generally reported higher self-concept beliefs than girls, but with a small magnitude of difference (Else-Quest, Hyde, & Linn, 2010). It is possible that boys and girls form their confidence in different ways. For example, for samples of secondary school students in the United States, current grades had stronger associations with students' science confidence expressed as self-efficacy beliefs for girls compared to boys (Britner, 2008; Britner & Pajares, 2006). Similarly, for a nationally-representative sample of secondary school students in the United States, mathematics grades had a larger positive influence on perceived mathematics ability for girls compared to boys; boys assessed their mathematical competence to be higher than girls for the same mathematics grades (Correll, 2001).

While many academic subjects have been considered in prior studies, gender differences in attitudes have indeed been observed for science in various countries (Nagy, Trautwein, Baumert, Köller, & Garrett, 2006; Simpkins, Davis-Kean, & Eccles, 2006; Taskinen, Schütte, & Prenzel, 2013). For example, for secondary school students in England, compared to girls, boys have expressed that science was easier, more interesting, more preferable than other subjects, and more relevant for careers, and boys also expressed higher aspirations to become a scientist or work in technology (Jenkins & Nelson, 2005). Differences have also been observed across specific science subjects. For example, for secondary school students in Germany, boys have reported higher self-concept beliefs in physics and chemistry while differences have been minimal for biology, regardless of whether students' attainment was controlled for or not (Jansen, Schroeders, & Lüdtke, 2014). Similarly, in England, secondary school boys have generally reported more-positive science self-concept beliefs than girls, and especially so for physics, but not for biology where no difference was observed (Hardy, 2014).

Essentially, various differences in attitudes across boys and girls, for example, may help explain why they report different intentions towards studying science. Differences in intentions across other observable groups of students may similarly follow from differences in attitudes. For example, families from particular backgrounds/ethnicities, such as Chinese or Indian, have often valued education and promoted professional or vocational careers for their children (Archer & Francis, 2006; Connor, Tyers, Modood, & Hillage, 2004; Strand, 2007; Torgerson, et al., 2008). Students from such backgrounds/ethnicities have accordingly often considered science and mathematics favourably, for example as indicators of success and as pre-requisites for following particular careers (Hernandez-Martinez, et al., 2008; Wong, 2012).

Differences in attitudes may follow from various influences. For example, students' perceptions of their classroom experiences can vary, including the support given or their teachers' expectations of them, which can in turn influence students' interest and confidence (Lazarides & Watt, 2015; Murphy & Whitelegg, 2006; Wang, 2012). Problematically, some teachers may also have lower expectations or perceive lower abilities in students with different backgrounds, regardless of their attainment (Campbell, 2015; Strand, 2007). For secondary school students, aspects of gendered stereotypes for academic subjects have also been observed, such as students believing that mathematics was a male domain and that boys were more likely to need mathematics skills for employment, to varying extents, in Canada (Plante, de la Sablonnière, Aronson, & Théorêt, 2013), France (Chatard, Guimond, & Selimbegovic, 2007), and Sweden (Brandell & Staberg, 2008). Students' conceptions of scientists have also often been highlighted to involve assumptions or stereotypes regarding gender and/or ethnicity/background (Archer, DeWitt, & Osborne, 2015; Carlone & Johnson, 2007; Carlone, Johnson, & Scott, 2015; Wong, 2015). As a specific example, for science students at university in the United States, stronger gender-science stereotypes associated with lower science identification and lower intentions to persist in science for women, but higher identification and intentions for men, although the difference was small (Cundiff, Vescio, Loken, & Lo, 2013).

Further research has also considered whether naturally-occurring clusters of students might hold similar or different beliefs, which may help characterise those students who may or may not intend to study science further, perhaps more so than considering students grouped by characteristics such as gender or background. Various studies of secondary school students in the United States have considered various attitudes and beliefs (such as interest, perceived utility, and self-concept beliefs), and have identified distinct clusters of students, each with broadly proportionate attitudes and beliefs such as these being generally high, moderate, or low in magnitude, although with various exceptions (Andersen & Chen, 2016; Andersen & Cross, 2014; Chow, Eccles, & Salmela-Aro, 2012; Conley, 2012; Simpkins & Davis-Kean, 2005). When further characteristics of students have been considered, boys have been more likely to be classified into clusters with generally high beliefs, while girls have been more likely to be classified into clusters with generally low beliefs (Chow, Eccles, & Salmela-Aro, 2012; Simpkins & Davis-Kean, 2005). When students' intentions or choices were also considered, students in clusters with consistently high beliefs were associated with studying more science and mathematics courses (Simpkins & Davis-Kean, 2005) and associated with higher aspirations to science careers (Chow, Eccles, & Salmela-Aro, 2012).

Any inferences and direct comparisons remain difficult, however, due to the differing factors considered and methods applied by the various studies. For example, considering Grade 9 secondary school students in the United States revealed clusters of students with broadly proportionate science self-efficacy beliefs (expressions of their confidence in their future abilities/attainment), personal value, utility value, and interest/enjoyment value, such as all factors being generally below-average (40% of the students), average (43%), or above-average (9%), but with one exception (Andersen & Chen, 2016). Specifically, one cluster (8% of students) exhibited above-average science self-efficacy and personal value, but average utility and moderately above-average interest (Andersen & Chen, 2016). Broadly similar profiles were also observed for mathematics (Andersen & Cross, 2014). Considering science self-concept (confidence in their current abilities/attainment), expectancy (expected future attainment and potential to become a good scientist), and value (interest and utility beliefs), for Grade 8 and 9 secondary school students in the United States highlighted that the majority of students (57%) formed a cluster with generally low responses to all the items, while a smaller proportion (12%) formed a cluster with generally high responses; another cluster (9%) reported generally high self-concept beliefs but low expectations, interest, and utility, while another cluster (21%) reported moderate self-concept, low expectations, high interest, and moderate utility (Aschbacher, Ing, & Tsai, 2014). Essentially, it remains difficult to infer whether distinct clusters with notably varying patterns of beliefs (such as holding high attitudes in one area but moderate or low attitudes in another) are exceptional, perhaps relating more to particular samples, or generalizable tendencies. Concurrently, while revealing distinct clusters of students is informative, for example in highlighting proportions of students who might plausibly continue to study science or not, it remains unclear what might cause students to hold particular patterns of beliefs (i.e. and be assigned to one cluster rather than another).

Few comparable studies appear to have been undertaken in England, although broad clusters of secondary school students (across ages 11-19) have been identified in relation to science (Institution of Mechanical Engineers, 2014). One cluster (29% of the sample) was found to be relatively focused on science, technology, engineering, and mathematics (STEM), generally from affluent backgrounds and also with links to STEM, with slightly more boys than girls (Institution of Mechanical Engineers, 2014). Another cluster (26%) was relatively focused on humanities, with less interest in STEM but not necessarily with lower confidence in their abilities, and who were engaged with their education, but generally from backgrounds without links to STEM (Institution of Mechanical Engineers, 2014). Another (17%) was relatively focused on vocational areas such as law, psychology, and business, but with less confidence and interest in STEM, and with more students from ethnic minority backgrounds (Institution of Mechanical Engineers, 2014). Another cluster (10%) covered those with interest in STEM but also interest in various other areas, but with low confidence and higher beliefs that STEM was difficult and only for the cleverest students, and with lower STEM aspirations (STEM was seen as not for 'people like them'), with more from ethnic minority backgrounds (Institution of Mechanical Engineers, 2014). The final cluster (18%) covered those who were less engaged in education in general, with lower confidence and fewer wider support networks, and with the lowest STEM aspirations, and who were generally from backgrounds without links to STEM (Institution of Mechanical Engineers, 2014). Nevertheless, the particular methodological details were not described in detail, and it perhaps remains unclear whether students would form similar or different clusters when considered at specific ages or over time.

In more general terms, recent research has proposed that groups of students in England, in Years 7 to 10, can be identified with varying magnitudes of 'science capital', considered as mainly the aggregate of someone's perceived utility of science, encouragement to continue studying science, access to science outside of school, and their perceptions of their parents' attitudes to science (Archer, Dawson, DeWitt, Seakins, & Wong, 2015). A minority of students (5%) were identified with high science capital, and such students were more likely to be male, South-Asian, with access to many cultural resources, and who were interested in studying science further, were confident in their abilities, and who were secure in their science identity (Archer, Dawson, DeWitt, Seakins, & Wong, 2015). The majority of students (68%) were identified with medium science capital while others were identified with low science capital (27%) (Archer, Dawson, DeWitt, Seakins, & Wong, 2015). Such groups rely on conceptual or theoretical distinctions and group boundaries, however, in contrast to clustering approaches where any cluster boundaries are empirically determined in order to maximise the fit to the data.

In even more general terms, various qualitative studies in the United States have highlighted that different students may hold different orientations towards science as a fundamental aspect of their personal identity, and/or embody or express a science identity in different ways (Carlone & Johnson, 2007; Carlone, Johnson, & Scott, 2015; Carlone, Webb, Archer, & Taylor, 2015; Tan, Calabrese Barton, Kang, & O'Neill, 2013). Similarly, qualitative studies in England have highlighted how different students may hold different profiles of beliefs, considered as varying combinations of science aspirations, attainment, interest, and other factors (Wong, 2016). Generally, such qualitative studies have often been relatively small in scale, however, and while these can highlight the plausibility of particular patterns and/or processes, they cannot necessarily highlight how prevalent they may be in wider samples.

In summary, different students may hold different attitudes, and patterns of various attitudes considered together. It remains unclear whether students' science attitudes broadly cohere in magnitude or whether other patterns may be observed (Andersen & Chen, 2016; Aschbacher, Ing, & Tsai, 2014). It also remains relatively unclear what might explain or cause students to hold particular patterns of attitudes and beliefs. Nevertheless, as before, considering various clusters of students appears to be a promising method.

Section 2.2.3: Differences in decision-making processes

Less research has considered whether different students have different approaches to considering their choices. Attitudes appear to be central to students' science choices, but different students might consider their choices in different ways, for example through considering different attitudes or wider factors to be more or less important or relevant to their decisions.

Some studies have extended considering gender differences in attitudes to consider gender differences in decision-making processes. For example, for students starting secondary school in Canada, girls reported higher intentions to continue with non-compulsory mathematics than boys, and these intentions were positively predicted by the students' utility value associated with mathematics for both girls and boys (Crombie, et al., 2005). For girls, mathematics intentions were also predicted by their mathematics self-concept beliefs, while for boys, intentions were also predicted by their mathematics attainment (Crombie, et al., 2005). Conversely, in the Netherlands, for students in early secondary school, the effect of mathematics attainment on mathematics choices was stronger for girls compared to boys, although at all levels of attainment boys were generally more inclined to study mathematics (Bosker & Dekkers, 1994). For secondary school students in Switzerland, the effect of science attainment on science choices was stronger for boys compared to girls, while the effect of interest was stronger for girls than boys (Aeschlimann, Herzog, & Makarova, 2016). Given the differences in considered factors across such studies, and their potentially contrasting results, it remains difficult to generalise any particular tendencies, other than that different students such as girls and boys may broadly consider their choices differently. This also relies on an assumption that predictive models indeed reflect processes of decision-making.

Further studies have considered more extensive arrays of attitudes and considered science directly. For example, for upper-secondary students in Norway, compared to boys, girls retrospectively reported that interest and enjoyment, fit to personal beliefs, and expectations of success were more important for selecting their subject area, while there were no gender differences for self-realisation (developing skills and being challenged), utility for university admission, and relative costs (Bøe, 2012). Concurrently, a subject's utility for university admission was considered more important, and interest and absence of costs were considered less important, for those studying science than for those studying humanities (Bøe, 2012). In the context of retrospectively describing important reasons for their A-Level subject choices, upper-secondary students in England have similarly cited utility (for studying the subject further at university, for being useful to future careers, and as being a ‘good subject’ to have studied) more for science and mathematics subjects, and cited interest and enjoyment more for humanities subjects (Vidal Rodeiro, 2007). In England, interviews with secondary school students have also suggested that girls and boys may hold different reasons for studying upper-secondary science: boys often reported ‘intrinsic’ reasons (e.g. because they enjoyed science, or wanted to understand the world) while girls often reported ‘extrinsic’ reasons (e.g. aiming towards careers or to help others), using such terms relatively broadly (Department for Children, Schools and Families, 2009). Particular sets of results may then appear similar or to contrast in some areas, but it remains unclear how this may reflect differences in samples or methods.

Other studies have highlighted differences in students’ approaches to choices, considered more broadly or indirectly. Interviews with secondary school students in England (followed from Years 9 to 11) have suggested that some students had already resolved on specific careers (whether in science or in other fields), some were partially resolved towards a specific career, some narrowed and refined their ideas over time, and some had changing, broad, or uncommitted career ideas (Cleaves, 2005). Those who did select non-compulsory science subjects reflected these broad trajectories: some students who chose science had already resolved or partially-resolved on a career in science (Cleaves, 2005). For those with broad or uncommitted ideas, science was sometimes studied as one subject within a broad mix, but the students were aware of the value of science for their subsequent options or careers; however, many who refined their ideas over time eliminated science as an option through actively selecting or focusing on other areas (Cleaves, 2005). This indirectly highlights that not studying science may follow from favouring other areas, rather than necessarily following from disfavouring science.

Students’ attainment may also indirectly influence their processes of considering their choices. For example, it is possible that some students may consider that they do not necessarily have a choice if they have lower attainment (i.e. contextual constraints of pre-requisite grades may eliminate choice for them), while students with higher attainment may need to consider further factors to actively make a choice between studying different subjects. For example, for secondary school students in England studying for GCSEs, expected attainment in mathematics appeared closely linked to intentions: most students predicted to attain grades of A* or A were considering or actually intending to continue with mathematics while those predicted to attain B or C reported much lower inclinations (Brown, Brown, & Bibby, 2008). The students' expressed reasons for not intending to continue with mathematics also appeared to differ depending on their expected attainment: for all predicted grades except for A*, students reported that they considered A-Level mathematics to be too difficult, and (to a lesser degree) that they did not enjoy or like mathematics; for those predicted to attain A* grades, however, the most frequently reported reason for not continuing mathematics was that mathematics was not necessary for their intended future degree or career (Brown, Brown, & Bibby, 2008). It remains unclear whether students' confidence in their abilities would similarly entail differences in decision-making processes, but it is plausible to assume that this may be the case. Students' expectations of their future capabilities, such as their expected attainment, have indeed been considered to be expressions of confidence, specifically self-efficacy beliefs (Bandura, 1997; Bong, 2001b).

Considered in general terms for students in England, students' backgrounds have also been proposed to influence their attitudes, specifically where various cultural aspects, family practices, and wider assumptions about science and scientists may help facilitate or limit students' potential identification with science (Archer, et al., 2012, 2013; Archer, DeWitt, & Wong, 2014). Such students would essentially be undertaking different processes of decision-making through implicitly considering various factors as being relevant or irrelevant to them, and/or considering what choices may be feasible or unfeasible. However, this does not appear to have been considered quantitatively.

In summary, few studies appear to have explicitly considered differences in students' decision-making. Studies have also applied quantitative approaches, such as comparing the predictive magnitudes of factors across girls and boys (Crombie, et al., 2005) or comparing self-reported reasons for choices (Bøe, 2012), and qualitative approaches such as broadly exploring the area through self-reported reasons via interviews (Cleaves, 2005). The fundamental points of note were that students' attitudes may have varying importance for different students (Bøe, 2012), and that attainment may entail differences in how choices are made (Brown, Brown, & Bibby, 2008). It appears plausible to hypothesise that students' confidence may similarly entail differences in how choices are made.

Section 2.3: An expectancy-value model of intentions and choices

The importance of students' attitudes and beliefs has been reflected through research in science education progressing from considering attitudes in generalised and aggregated forms (Osborne, Simon, & Collins, 2003) to conceptualising and considering specific attitudes and motivational beliefs (Regan & DeWitt, 2015) and by increasingly applying motivational theories as interpretative and explanatory perspectives (Bøe & Henriksen, 2015). For example, recent research has often applied the expectancy-value model of motivated behavioural choices (Eccles, 2009; Eccles, et al., 1983; Wigfield & Eccles, 2000) to gain wider insight into students' intentions and choices (Bøe, 2012; Bøe & Henriksen, 2013, 2015; Bøe, Henriksen, Lyons, & Schreiner, 2011; Wang & Degol, 2013).

The expectancy-value model aims to provide a relatively comprehensive theoretical framework for studying the psychological and contextual factors underlying motivation, attainment, and choices within education (Eccles, 2009; Eccles, et al., 1983; Wigfield & Eccles, 2000). Educational and career choices are assumed to closely relate to someone's expectations for success and the value attached to the various options considered as available; these various beliefs are assumed to be influenced by cultural factors, someone's family and background, someone's context and environment, and their affective and other reactions from earlier experiences (Eccles, 2009). The model emphasises subjectively-perceived values and interpretations, recognising that students may have limited information, understand information differently, and may not necessarily follow a 'rational' decision-making process (Bøe, Henriksen, Lyons, & Schreiner, 2011; Eccles, 2009).

Within the model, specific attitudes and beliefs associated with activities and wider areas have been considered, and formally termed ‘subjective task values’ (Eccles, 2009). These have been conceptualised as interest value (intrinsic interest and enjoyment in doing something), utility value (extrinsic usefulness as doing something as a means to gain wider benefits or outcomes), attainment value (the importance or value an activity has through affirming a personal or collective identity), and cost value (such as financial and emotional costs as well as time and other aspects) (Eccles, 2009). (This thesis uses ‘personal value’ to refer to ‘attainment value’ for intuitive clarity, for example to avoid confusion with higher attainment (performance or achievement) being valued.) Additionally, the model has separated someone’s confidence in their personal abilities (akin to self-concept beliefs) from their expectations for success (akin to self-efficacy beliefs), although the two aspects have been proposed to influence each other and have sometimes been measured as if they were the same (Eccles, 2009; Wigfield & Eccles, 2000).

These various elements have been assumed to closely, and reciprocally, associate (Eccles, 2009). In accordance with these assumptions, students’ attitudes, attainment, and beliefs of their own abilities have associated in various ways. For example, across primary/elementary school and secondary school in the United States, students’ interest and self-concept beliefs closely linked, and to a greater extent than the link between attainment and self-concept and the link between attainment and interest (Denissen, Zarrett, & Eccles, 2007). Similarly, again in the United States, students’ self-concept beliefs and interest have both reciprocally associated with attainment at upper-secondary school when measured from Grade 10 to five years after secondary school graduation (Guo, Marsh, Morin, Parker, & Kaur, 2015). Other close associations have been observed in various other studies. For example, for Grade 9 students in Germany for science, interest and attainment (Jansen, Lüdtke, & Ulrich, 2016), and interest and perceived utility (Taskinen, Schütte, & Prenzel, 2013), have closely associated. Similarly, for Grade 9/10 students in Norway for mathematics, interest and perceived utility, and interest and effort, have associated (Federici & Skaalvik, 2014).

The expectancy-value model broadly assumes that students’ context and background influences their subjective values and beliefs, which then influence their intentions and choices (Eccles, 2009). These assumptions have been supported by various research studies. For example, parents' and teachers' expectations and attributions may influence students' beliefs, such as through parents and teachers attributing success due to ability in boys but due to effort in girls (Espinoza, Arêas da Luz Fontes, & Arms-Chavez, 2014; Gunderson, Ramirez, Levine, & Beilock, 2012). For secondary school students in Australia, boys and girls differentially perceived the beliefs of their mathematics teachers: compared to boys, girls believed that their teachers expected them to do less well, but believed that their teacher thought that mathematics had higher status compared to other subjects (Lazarides & Watt, 2015). For secondary school students in the United States, students' classroom experiences, including their beliefs about their teachers' expectations of the student doing well, the level of social support, and the extent of promoting understanding and discussing the meaning of problems and issues, predicted the students' aggregated perceived importance, utility, and interest (Wang, 2012). The students' course enrolment and career aspirations in mathematics were then positively predicted by their expectancies (mixing self-concept and self-efficacy items) and their values (aggregated importance, utility, and interest), but were not directly predicted by their classroom experiences (Wang, 2012).

Theoretical models may ideally help ensure that any inherent elements are similarly conceptualised across different studies, although different results may still follow from different interpretations or implementations of these elements (Bong, 1996). For example, the expectancy-value model separated someone's confidence in their personal abilities from their expectations for success, although it was subsequently recognised that various research studies had used the terms and/or different measurement items interchangeably (Eccles, 2009; Wigfield & Eccles, 2000). Similarly, various studies have considered students' values aggregated together (Chow, Eccles, & Salmela-Aro, 2012; Wang, 2012) or as discrete factors (Musu-Gillette, Wigfield, Harring, & Eccles, 2015). Even when studies have considered students' values as an aggregate, each of the four theorised elements has not always been covered; for example, sometimes only interest and utility have been considered (Wang, 2012).

On a wider level, the expectancy-value model has also been contextualised with reference to personal identity (Eccles, 2009). Someone's perceptions of various cultural and social aspects may influence how they orient their personal identity against various roles, relationships, and expectations; various actions may be valued or promoted through helping to enact a particular role or identity (Eccles, 2009). Essentially, the expectancy-value model has been interpreted to inherently link students' aspirations for the future with who they want to become in the future (Eccles, 2009). Students' identities have been increasingly considered within various research fields, and have been defined and explored in various ways (Côté, 2006; Gee, 2000; Sfard & Prusak, 2005). Within science education, considering students' identities has encompassed numerous aspects such as: students' current beliefs of themselves and their abilities; students' beliefs of how other people view them and wider social and cultural expectations and influences; and students' beliefs of what they want to do and become in the future (Archer, et al., 2010; Aschbacher, Li, & Roth, 2010). A complementary perspective has considered that enacting or embodying a particular identity may require specific knowledge and skills (or 'competences') in undertaking relevant social practices (or 'performances'), which are recognised and observed by others and also by the person themselves (Carlone & Johnson, 2007). These elements highlight the importance of personal skills and confidence, including self-reflection and recognising such skills.

In summary, the expectancy-value model proposes a plausible framework for students' intentions and choices, highlights the importance of students' confidence (which may be conceptualised and expressed as akin to self-concept and self-efficacy beliefs), and has been supported through various research studies (Bøe & Henriksen, 2015; Wang & Degol, 2013). Practically, the model provides a basic list of factors to consider, which can then be extended as necessary. While the various factors within the expectancy-value model have been historically measured slightly differently or included or omitted (Eccles, 2009; Wigfield & Eccles, 2000), specific models of educational choices may help ensure that the considered factors broadly reflect historically-accepted and/or contemporary theory and conceptualisations (Cohen, Manion, & Morrison, 2007; Messick, 1995).

Section 3: Students' confidence

Students' confidence, which is used here as an intuitive term for brevity and inclusivity, has been conceptualised in various ways within educational and motivational research, most commonly as 'self-concept' and 'self-efficacy' beliefs (Bandura, 1977; Bong & Clark, 1999; Bong & Skaalvik, 2003). Someone's current confidence ('self-concept') broadly considers their interpretations of their historic and current attainment experiences, such as gaining particular grades or accomplishing difficult work, often through evaluations of whether they think that they are 'doing well' or are 'good' at a subject (Bong & Skaalvik, 2003; Shavelson, Hubner, & Stanton, 1976). Alternately, someone's confidence in their future capacities or capabilities ('self-efficacy') considers their evaluative beliefs about specific events or contexts, such as their confidence in being able to gain a particular examination grade or in being able to successfully undertake a particular type of task (Bandura, 1977; Bong & Skaalvik, 2003).

Section 3.1 briefly considers these common conceptualisations of students' confidence, self-efficacy and self-concept, and how they may be influenced; various factors, other than attainment, have been theoretically proposed and empirically found to influence students' confidence, which may potentially introduce under-confidence or over-confidence biases. **Section 3.2** then covers prior research that has directly considered confidence accuracy/biases. A plausible conceptual and theoretical rationale for the importance of accurate confidence is then considered in **Section 3.3**, specifically through the idea of self-regulation. Additionally, given the focus on confidence accuracy/bias throughout this thesis, and given that it needed to be calculated within the analysis, **Section 3.4** discusses some of the inherent uncertainty involved in attempting to measure confidence accuracy/bias, and how prior research has calculated indicators of accuracy/bias.

Section 3.1: Confidence as self-efficacy and self-concept beliefs

Confidence conceptualised as self-efficacy forms an integral aspect of social-cognitive theory (Bandura, 1997). Social-cognitive theory essentially proposes that individual behaviour is not necessarily entirely autonomous, nor necessarily entirely deterministic; instead, behaviour, motivations and beliefs (including confidence), and wider contextual influences are assumed to each reciprocally influence one another (Bandura, 1989).

Self-efficacy represents someone's beliefs in their capabilities to successfully undertake an action (Bandura, 1977). Self-efficacy was considered to be distinct from outcome expectations (estimates that a particular action would then lead to a particular outcome) and was theorised to be specific to tasks and domains (Bandura, 1977, 1986, 1997). Subsequent studies have accordingly confirmed that students' self-efficacy beliefs have been specific to various areas, but have nevertheless often associated to some degree across academic subjects and across more-specific to more-generalised measures (Bong, 1997, 2001a, 2002, 2004).

High self-efficacy beliefs may be motivational and facilitate people to surpass their normal performance, while low self-efficacy beliefs may be limiting and ensure that some actions are not even attempted (Bandura, 1997). In accordance with these theoretical assumptions, higher self-efficacy has indeed been associated with various motivational approaches, such as aiming to learn and master academic work (Jiang, Song, Lee, & Bong, 2014; Phillips & Gully, 1997), with persistence (Multon, Brown, & Lent, 1991; Skaalvik, Federici, & Klassen, 2015), and with students' self-regulation for their learning (Usher & Pajares, 2008a; Zimmerman & Schunk, 2011). Self-efficacy has also strongly predicted students' attainment (Britner, 2008; Britner & Pajares, 2006; Multon, Brown, & Lent, 1991) and has associated with their studying intentions (Bong, 2001b; Pajares & Miller, 1995; Stevens, Wang, Olivárez, & Hamman, 2007).

Self-efficacy beliefs were theorised to follow from particular sources or antecedents (Bandura, 1977, 1997): 'enactive mastery experiences' (i.e. prior experiences, whether successes or failures); 'vicarious experiences' (i.e. seeing others successfully perform, which was assumed to be more influential when there were no absolute measures of adequacy and when there was a perceived similarity with the person performing); ‘verbal persuasions’ (i.e. persuasive and evaluative feedback from significant others, which was assumed to be more influential when the person was considered knowledgeable and credible, and the information was considered realistic, but which was assumed to be easily outweighed by disconfirming mastery experiences); and ‘physiological reactions’ (i.e. physical and emotional responses such as anxiety).

Mastery experiences have generally been found to be the most influential from these four sources of self-efficacy beliefs, while the influences of the others have varied across studies (Britner, 2008; Britner & Pajares, 2006; Lopez, Lent, Brown, & Gore, 1997; Usher & Pajares, 2008b, 2009). Correlations between the sources of self-efficacy have been typically found, since those gaining successful experiences likely receive praise for them and experience positive feelings (Usher & Pajares, 2008b). Relatively similar findings have also been observed through qualitative explorations of influences on self-efficacy (Butz & Usher, 2015; Zeldin & Pajares, 2000; Zeldin, Britner, & Pajares, 2008) and through interventions that have targeted self-efficacy (van Dinther, Dochy, & Segers, 2011).

Further studies have highlighted that different students may be influenced in different ways. For example, the four theorised sources have differentially predicted self-efficacy for boys and girls (Britner, 2008; Britner & Pajares, 2006; Lent, Lopez, & Bieschke, 1991). Additionally, distinct clusters of students have been identified who reported varying magnitudes of the four theorised sources (Chen & Usher, 2013). Such differences may then help explain differences in students’ confidence, and/or confidence biases towards under-confidence or over-confidence (although such studies have not explicitly made the connection to confidence biases). However, prior studies have often focused on determining the relative magnitudes of the four theorised sources (whether via reported expressions or via coefficients in predictive models), with less focus on considering any other potential antecedents/influences or wider implications.

Confidence conceptualised as self-concept evolved from general psychological measures (such as self-esteem) rather than within a motivational theory (Shavelson, Hubner, & Stanton, 1976). Self-concept was originally conceptualised as someone's perceptions of their self, formed through experiences and interactions with and within their environment (Shavelson, Hubner, & Stanton, 1976). Concurrently, self-concept was ascribed various characteristics, such as being structured, hierarchical, and being both descriptive and evaluative, where evaluations could be made against absolute standards or relative standards such as peers (Shavelson, Hubner, & Stanton, 1976). However, someone's perceptions of their self are many and varied, and it perhaps remained unclear regarding what, exactly, should be measured. Initially, self-concept encompassed confidence together with interest and enjoyment (Marsh & Shavelson, 1985). Subsequently, self-concept became increasingly focused on someone's beliefs of their academic ability (i.e. evaluations of 'being good' or 'doing well'), and students' affective responses were then considered separately (Arens, Seeshing Yeung, Craven, & Hasselhorn, 2011; Marsh, Craven, & Debus, 1999). Contemporary conceptions of self-concept beliefs do not, therefore, consider someone's overall sense or idea of their 'self' as it would be considered through 'science identity' (Archer, et al., 2010; Aschbacher, Li, & Roth, 2010; Carlone & Johnson, 2007) or through wider ideas of identity (Côté, 2006; Gee, 2000; Sfard & Prusak, 2005).

Early research helped establish academic subject-specific self-concept beliefs that also generalised into wider factors such as verbal/language self-concept beliefs and mathematical/quantitative self-concept beliefs (Marsh, 1990; Marsh & Shavelson, 1985; Marsh, Byrne, & Shavelson, 1988). Early studies also found that students' mathematics and languages attainment generally positively associated, while mathematics and languages self-concept beliefs generally had very low or no associations, when the four factors were modelled together (Marsh, 1986b; Marsh, Byrne, & Shavelson, 1988). Additionally, someone's self-concept beliefs were positively predicted by their own academic attainment but negatively predicted by group-average attainment when these three factors were modelled together (Marsh, 1987; Marsh & Parker, 1984).

Results from these early studies were interpreted to mean that students formed their self-concept beliefs, to some extent, in reference to or in contrast with their peers (Marsh, 1987; Marsh & Parker, 1984). The results were subsequently replicated across England (Nagengast & Marsh,

2011) and various other countries via data from international studies (Marsh & Hau, 2003; Marsh, Abduljabbar, et al., 2015; Seaton, et al., 2008; Seaton, Marsh, & Craven, 2009). Students were concurrently assumed to form their self-concept beliefs in reference to or through contrasting their attainment in different subjects (Marsh, 1986b; Marsh, Byrne, & Shavelson, 1988). Similarly, these results were subsequently replicated using data from various international studies (Chiu, 2008, 2012; Marsh, et al., 2014; Marsh, Lüdtke, et al., 2015; Möller & Köller, 2001; Möller & Marsh, 2013; Möller, Pohlmann, Köller, & Marsh, 2009). Such results, however, have generally not been observed for self-efficacy (Bong, 1998; Möller, Pohlmann, Köller, & Marsh, 2009; Skaalvik & Rankin, 1990, 1995).

Outside of replication studies, self-concept studies have produced more variable results. Longitudinal studies have variously found some evidence (Marsh & Köller, 2004; Niepel, Brunner, & Preckel, 2014b) or no evidence (Chen, Yeh, Hwang, & Lin, 2013; Möller, Retelsdorf, Köller, & Marsh, 2011; Parker, Marsh, Morin, Seaton, & Van Zanden, 2015) for self-concept beliefs being influenced by contrasting attainment across different subjects. Students' self-concept beliefs have also positively associated within similar academic areas (such as across quantitative subjects including mathematics and science) (Jansen, Schroeders, Lüdtke, & Marsh, 2015; Marsh, et al., 2014; Marsh, Lüdtke, et al., 2015), and when students had similar attainment across different subjects (Rost, Sparfeldt, Dickhauser, & Schilling, 2005). Self-concept research has also often focused on replication of specific models, but without exploring how and why any assumed effects such as peer-comparisons occur. Numerous factors may be relevant, such as who students compare themselves against and the specific practices of grouping students within schools (Chiu, et al., 2008; Collins, 1996; Huguet, Dumas, Monteil, & Genestoux, 2001; Ireson & Hallam, 2009; Lubbers, Kuyper, & van der Werf, 2009; Wehrens, Kuyper, Dijkstra, Buunk, & van der Werf, 2010). Wider psychological research has also suggested that peer-comparisons may be inherently problematic to explore: in some situations, students may base their peer-comparisons mainly on assessments of themselves rather than on assessments of others, so that a comparison may not necessarily take place (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Klar & Giladi, 1999; Kruger, 1999). Broadly, some uncertainty remains into what might influence self-concept beliefs: self-concept research has almost exclusively been quantitative, and has not included interviews or free-responses to allow students to report what influences their confidence beliefs, in contrast to various studies into self-efficacy (Butz & Usher, 2015; Zeldin & Pajares, 2000; Zeldin, Britner, & Pajares, 2008).

The motivational implications of self-concept beliefs have been explored for various areas. For example, across various samples and countries, prior attainment has predicted subsequent self-concept beliefs, controlling for prior self-concept beliefs, and concurrently prior self-concept has predicted subsequent attainment, controlling for prior attainment (Huang, 2011b; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; Möller, Retelsdorf, Köller, & Marsh, 2011; Niepel, Brunner, & Preckel, 2014a; Preckel, Niepel, Schneider, & Brunner, 2013; Seaton, Parker, Marsh, Craven, & Yeung, 2014; Skaalvik & Hagtvet, 1990). Self-concept may then have a motivational role, similar to self-efficacy, although it perhaps remains unclear why. Various studies have revealed close associations between self-concept beliefs and interest over time (Denissen, Zarrett, & Eccles, 2007; Guay, Ratelle, Roy, & Litalien, 2010; Guo, Marsh, Morin, Parker, & Kaur, 2015; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; Viljaranta, Tolvanen, Aunola, & Nurmi, 2014). It is perhaps then plausible to infer that, similarly to self-efficacy, higher self-concept may appear motivational through perhaps entailing higher interest and engagement.

In summary, self-efficacy beliefs, and perhaps also self-concept beliefs, may be motivational, and associate with beneficial outcomes within education. Self-efficacy and self-concept beliefs may be influenced by broadly similar areas within education; for example, peers may provide evidence that students can similarly accomplish tasks and so influence self-efficacy beliefs, and/or peers may partially provide evaluative standards for what ‘being good at science’ means and so influence self-concept beliefs. These various influences on students’ confidence, other than their attainment, may potentially introduce biases towards under-confidence or over-confidence. Additionally, similarities in self-efficacy and self-concept beliefs from more-specific to more-generalised beliefs, and across areas or subjects, may suggest that any biases towards under-confidence and over-confidence may be similarly reflected across degrees of generalisation and/or across academic subjects.

Section 3.2: Confidence accuracy and biases

The motivational implications of higher self-efficacy and/or self-concept beliefs suggest that higher confidence may be beneficial in itself (Bandura, 1997). It is possible that benefits then occur even if someone is over-confident, although the area remains relatively unclear.

On a wider level, views have also differed as to whether over-confidence, self-enhancement, self-serving biases, positivity biases, positive illusions, and similar concepts are essentially universal and potentially 'beneficial' to everyone (Cai, et al., 2011; Gaertner, Sedikides, & Chang, 2008; Hepper, Gramzow, & Sedikides, 2010; Hepper, Sedikides, & Cai, 2013; Sedikides, Gaertner, & Toguchi, 2003), or whether the potential benefits of any biases may be relative to particular countries (Heine, 2005; Heine & Hamamura, 2007; Heine, Lehman, Markus, & Kitayama, 1999; Kitayama, Markus, Matsumoto, & Norasakkunkit, 1997). The issue may be complicated through differences in some biases across countries being dependent, to some extent, on the methods used to reveal and measure them (Church, et al., 2014; Krueger & Wright, 2010; Mezulis, Abramson, Hyde, & Hankin, 2004). Regardless, for students within England (or any country considered alone), it is perhaps more relevant to explore whether they are under-confident or over-confident compared to other students in England, or when compared against absolute standards. Accordingly, international differences (in the sense of universal/relative biases) are not focused on when considering prior literature, and within the overall research design for the thesis.

The following sections review various studies into students' accuracy/bias, in order to consider the different ways in which under-confidence, accuracy, or over-confidence might be considered beneficial or detrimental, such as associating with higher or lower attainment (**Section 3.2.1**) and/or higher or lower attitudes and other aspects relevant to education (**Section 3.2.2**). Less research has explored how prevalent different degrees of accuracy/bias may be across samples of students (**Section 3.2.3**); however, such studies have potentially highlighted the benefit of exploring confidence accuracy/bias via cluster analysis.

Section 3.2.1: Accuracy/bias and attainment

Many studies have explored how students' accuracy/bias associates with their attainment, which is often considered to be fundamentally beneficial within educational systems, although results have varied. For example, studies of secondary school students have, via different approaches, samples, and measured factors, variously associated under-confidence with higher attainment (Chiu & Klassen, 2010), accurately-evaluated beliefs with higher attainment (Chen, 2003), and also over-confidence with higher attainment (Dupeyrat, Escribe, Huet, & Régner, 2011). Perhaps unhelpfully, each case could then be considered to be 'beneficial', from a particular perspective and given a particular context and methodology.

Further studies can nevertheless provide insights into the area. For example, a study of students at the end of primary school in the Netherlands (Grade 5, age 10-11) showed that boys performed higher than girls on arithmetic mathematics tasks, and were more confident in their ability to find an adequate solution; while boys tended to over-estimate their performance (showing over-confidence) compared to girls, the girls were not necessarily under-confident (Boekaerts & Rozendaal, 2010). The students generally over-estimated their performance when measured after the tasks than before the tasks, although this varied by the type of mathematics task: accuracy was higher after completing arithmetic tasks, suggesting that the students were able to evaluate their problem-solving process and make an enhanced assessment of their performance, although the opposite occurred for applied tasks that required interpretation of the problem (Boekaerts & Rozendaal, 2010). For secondary school students in the United States (age 15), students' retrospective estimates of their performance were more accurate (although still slightly over-confident) compared to their predictions made in advance, although these did vary across academic subjects (Erickson & Heit, 2015).

For students in early secondary school in the United States (Grade 7, age 12-13), accuracy in confidence associated with higher attainment, both directly and indirectly via self-efficacy beliefs (Chen, 2003). For similar students, accuracy decreased and over-confidence increased with the difficulty of the tasks (Chen & Zimmerman, 2007). Similarly, again for secondary school students in the United States considered over time (Grades 5-8, age 10-14), accuracy was higher for easier rather than harder questions, for boys, and for those with higher performance (Rinne & Mazzocco, 2014). Additionally, higher accuracy predicted higher subsequent gains in mathematics performance, controlling for the students' concurrent performance (Rinne & Mazzocco, 2014). For all secondary school students (age 15) surveyed by PISA 2000, which measured students' self-concept beliefs in reading (Chiu & Klassen, 2009) and in mathematics (Chiu & Klassen, 2010), those who were over-confident in their self-concept beliefs were more likely to have scores below their country mean, while under-confident students were more likely to score above their country mean (Chiu & Klassen, 2009, 2010).

Studies with undergraduate students, most undertaken in the United States with psychology students, have usually highlighted higher accuracy but slight under-confidence in higher-performing students and lower accuracy and over-confidence in lower-performing students (Ackerman & Wolman, 2007; Bol, Hacker, O'Shea, & Allen, 2005; Hacker, Bol, Horgan, & Rakow, 2000; Maki, Shields, Wheeler, & Zacchilli, 2005). Similarly, undergraduate students in England generally over-estimated their mental arithmetic test scores, although higher-performing students had smaller biases (Chevalier, Gibbons, Thorpe, Snelle, & Hoskins, 2009). Such results have been observed since some of the earliest studies into accuracy/bias: historically, university students with lower performance were seen to be over-confident while students with higher performance were seen to be slightly under-confident (Lichtenstein & Fischhoff, 1977). Students were generally over-confident, but showed under-confidence on easier test items (Lichtenstein & Fischhoff, 1977). Similarly, and more recently, undergraduate students in the United States also over-estimated their scores when a test was difficult but not when it was easy (Krueger & Mueller, 2002).

In general terms, psychological studies with undergraduate students have applied various methods. Similar results have broadly been seen when undergraduate students have made relative estimates of their performance, such as estimating their percentile ranking against the study sample or against their peers in general (Kruger & Dunning, 1999), and also when students made absolute estimates, such as estimating the numbers of test items forecasted or retrospectively considered to be correct (Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008). However, higher-performing students could be variously under-confident or accurate when absolute estimates were used (Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008). Nevertheless, others have suggested that various effects, including statistical or methodological aspects, may have contributed to the varying results (Ackerman, Beier, & Bowen, 2002; Krajc & Ortmann, 2008; Krueger & Mueller, 2002).

Specific studies of undergraduate students in the United States have shown that over-evaluating their academic abilities was associated with lower subsequent course grades (Gramzow, Elliot, Asher, & McGregor, 2003; Robins & Beer, 2001). Further studies highlighted that similar students generally over-reported their current attainment (their Grade Point Average, GPA) and their prior attainment (their Scholastic Aptitude Test, SAT, scores) (Gramzow & Willard, 2006). Over-reporting their GPA was positively predicted by believing themselves to be better than other students on a range of personality measures (which may or may not have been accurate and could not be determined), while over-reporting their SAT scores was positively predicted by being an older student (Gramzow & Willard, 2006). The pattern was interpreted as highlighting that self-enhancement contributed to motivational biases in reporting current performance (which reflected current aims and goals) but was less relevant to biases in reporting prior performance (which reflected prior aims) (Gramzow & Willard, 2006). Reconstructive memory processes were suggested to contribute to the increased bias in reporting prior performance in older students, as older information may become less accessible over time and may increasingly be influenced by tendencies to bias reports (Gramzow & Willard, 2006; Willard & Gramzow, 2008). Further work has highlighted that any costs and benefits of over-confidence may depend on any underlying motivation: exaggeration of students' GPA associated with their reported need for achievement (e.g. enjoying challenging tasks and not necessarily working only because they were required to) and improvement in their subsequent academic attainment (Willard & Gramzow, 2009). In that situation, exaggeration may have reflected the projection of positive goals onto someone's responses (Willard & Gramzow, 2009).

Further studies with undergraduate students, again mostly occurring in the United States, have explored memory and accuracy/biases (considering meta-cognition or 'meta-memory' via 'judgements of learning'). Generally, specific methods have been applied, where students memorise pairs of words and report their confidence in subsequently recalling the second word when given the first word (i.e. providing their 'judgment of learning'), which is then compared against their subsequent test performance (Finn & Metcalfe, 2007; Koriat, 1997). Such studies have broadly highlighted that students were often over-confident prior to undertaking a test, but were under-confident on subsequent testing occasions (Finn & Metcalfe, 2007, 2008; Koriat, 1997; Koriat, Sheffer, & Ma'ayan, 2002). Such studies have also been undertaken with younger students, but highlighting slightly different results. For example, for primary school students in the United States (Grade 3 and 5, age 8 and 10), students remained over-confident in their predictive judgements of being able to remember definitions, across three occasions of studying, predicting, and testing (Finn & Metcalfe, 2014). The students were also very accurate for items that were correct but less accurate for items that were incorrect, essentially reflecting a higher 'false positive' reporting rate (Finn & Metcalfe, 2014). For primary school students in Switzerland (age 6), under-estimating students performed higher than over-estimating students and showed higher discrimination in their confidence judgments when they were incorrect (i.e. accurately reporting low confidence when they were incorrect) compared to other patterns (Destan & Roebers, 2015). The students showed similar discrimination when they were correct (i.e. reporting high confidence when they were correct) regardless of whether they were under-confident, accurate, or over-confident (Destan & Roebers, 2015). Such studies begin to suggest that under-confidence and over-confidence may follow from different processes, such as over-confidence perhaps following from students not recognising when answers might be wrong.

In summary, higher attainment is intuitively beneficial within education. While results have varied, it seems plausible to infer that accurate beliefs broadly associate with higher attainment (Chen, 2003). However, the area is potentially complicated since accuracy may be higher for easier tasks (Rinne & Mazzocco, 2014) and over-confidence may increase with the difficulty of the task (Chen & Zimmerman, 2007). It also remains unclear whether higher accuracy simply reflects higher attainment in some way or whether higher accuracy facilitates higher attainment in some way (Rinne & Mazzocco, 2014).

Section 3.2.2: Accuracy/bias and attitudes and beliefs

Various studies have explored the associations between accuracy/bias and students' wider attitudes and motivational beliefs, some of which may be (indirectly) considered beneficial within education, and/or relevant to students' studying choices.

For example, a study of Canadian primary school students (Grade 3 and 5, age 9 and 11) showed that under-confident students (via self-beliefs of general ability compared to a general mental ability test) reported lower intrinsic motivation for mathematics, pride in their results, and attitude to effort (i.e. being less likely to apply effort as a means to succeed), and gained lower attainment in mathematics, compared to accurate and over-confident students (Bouffard, Boisvert, & Vezeau, 2003). For primary school students in the United States (Grade 3, age 8-9), confidence biases were considered through comparing students' general academic self-concept beliefs against teachers' judgments of their academic attainment; those with accurately-high self-concept beliefs had higher reading and mathematics attainment, and higher social skills, than those with over-confident or accurately-low beliefs (Gresham, Lane, MacMillan, Bocian, & Ward, 2000). Those with accurately-high and over-confident beliefs similarly reported high perceived importance for their academic work (Gresham, Lane, MacMillan, Bocian, & Ward, 2000). Curiously, insufficient numbers of students could be classified into an under-confident group for that to be considered (Gresham, Lane, MacMillan, Bocian, & Ward, 2000).

For secondary school students in Greece (Grade 9 and 10, age 15-16), over-confident students (via their self-efficacy compared with their attainment) reported higher interest in the subject than under-confident students for both mathematics and languages; accurately-evaluating students reported more interest in mathematics compared to over-confident students, and reported the same interest as over-confident students for languages (Gonida & Leondari, 2011). No group differences were found for persistence, mastery goals (focusing on learning/mastering studying), and performance-approach goals (focusing on out-performing peers), for both mathematics and languages considered separately (Gonida & Leondari, 2011). However, when considering students who were over-confident, accurate, or under-confident in both mathematics and languages, over-confident students reported higher performance-approach goals (focusing on out-performing peers) and higher performance-avoidance goals (focusing on avoiding looking worse than peers) compared to the other groups; over-confident students and accurate students reported similar mastery goals, higher than under-confident students (Gonida & Leondari, 2011). For secondary school students in France (Grade 8 and 9, age 13-16), considering their perceived relative ability to other students compared against their relative attainment, over-evaluating students in mathematics reported higher performance-approach goals and made more progress during the year than under-evaluating students, while the over-evaluating and accurately-evaluating groups did not significantly differ (Dupeyrat, Escribe, Huet, & Régner, 2011). Alternately, for secondary school students in Austria (Grade 7, age 14), neither mastery goal orientations nor performance-avoidance goal orientations for school studying were predicted by an accuracy/bias indicator (self-estimated attainment compared against actual attainment) for German language and for mathematics (Schwab & Hessels, 2015).

For undergraduate students in the United States, exaggerating their attainment (GPA) positively associated with a performance-approach orientation but had no association with a performance-avoidance orientation or with a mastery orientation (Willard & Gramzow, 2009). For similar university students in Germany and the Netherlands, those with high performance-approach goals over-estimated their intelligence test scores and those with high performance-avoidance under-estimated, while mastery goals did not associate with over-estimating or under-estimating (Bipp, Steinmayr, & Spinath, 2012). University students in the United States who expressed a fixed/entity view of intelligence/ability tended towards over-confidence in estimating their test-scores compared to those who expressed an incremental/changeable view, who were more accurate (Ehrlinger, Mitchum, & Dweck, 2016).

Few studies have considered students' intentions or choices and confidence accuracy/bias. As one example, for secondary school students in Australia (Year 9 and 11, age 14-16), over-evaluation of their mathematics self-concept beliefs (compared to their teachers' perceptions) was associated with higher examination performance and greater intentions to engage with mathematics in the future (Martin & Debus, 1998). More recently in England, research has started to explore the implications of confidence accuracy/bias on students' intentions, considering accuracy/bias using paired tasks and confidence-ratings (Sheldrake, Mujtaba, & Reiss, 2014). For Year 8 students, under-confident students reported lower mathematics attitudes than accurate and over-confident students, including for their interest and perceived utility; at Year 10, however, over-confident students reported the lowest intentions to study mathematics further (Sheldrake, Mujtaba, & Reiss, 2014).

In summary, it is possible to infer that over-confidence may associate with higher interest and performance orientations (i.e. motivational tendencies to try to perform higher than peers), although results have varied across studies. It remains unclear whether perceptions of the utility of subjects and other attitudes or beliefs also differ across degrees of accuracy/bias.

Section 3.2.3: Accuracy/bias across clusters of students

Research has started to consider how prevalent different degrees of accuracy/bias may be, through defining and considering clusters of students with similar biases, although results have varied. Few studies have considered similar samples and/or applied similar approaches.

For example, one study of Canadian students from Grade 3/4 (age 9/10) to Grade 7/8 (aged 13/14) highlighted that most students had a stable degree of accuracy/bias over time (when comparing their general perceptions of ability against intelligence test scores): the majority (75% of the sample) showed a stable but moderate over-estimating bias, while a smaller cluster (15%) showed a stable but highly over-estimating bias; a minority (6%) showed an under-estimating bias that became even more under-estimating over time, while another minority (4%) showed a highly under-estimating bias that changed to a moderately under-estimating bias over time (Bouffard, Vezeau, Roy, & Lengelé, 2011). In comparison, clusters of students in primary school in Finland (Grade 1-2, age 7-8) were identified using their mathematics self-concept beliefs and their attainment: half of the students (49%) were over-confident with high self-concept but low attainment, a third (33%) were accurate with high self-concept and high attainment, and the remainder (18%) were accurate with low self-concept and low attainment (Rytkönen, Aunola, & Nurmi, 2007). Cluster membership was relatively stable over time, although the proportion of students in the over-confident group slightly decreased over two years while the proportion of students in the accurate clusters slightly increased (Rytkönen, Aunola, & Nurmi, 2007). Differences in results might simply reflect different studies measuring different students and areas, although it is possible that tendencies towards accuracy/bias may be relatively stable over time (although difficult to determine given limited research). Psychological studies with university students have also suggested stable bias over time (Kelemen, Frost, & Weaver, 2000).

Clusters of secondary school students in Spain (Grade 10-11, age 15-16) were identified when considering their mathematics self-concept and attainment (Sáinz & Upadyaya, 2016). Clusters formed for those with accurately-high self-concept beliefs (36% of the students), accurately-low beliefs (22%), over-confident beliefs (20%), and under-confident beliefs (21%) (Sáinz & Upadyaya, 2016). The students' cluster membership remained relatively constant between Grade 10 and Grade 11, although the proportion of students in the accurately-high cluster slightly decreased at

Grade 11, and the proportions of students in the other clusters increased accordingly (Sáinz & Upadyaya, 2016). Boys were more likely than girls to be assigned to the accurately-high (at Grade 10) or over-confident clusters (at Grade 10 and Grade 11) than the under-confident cluster (Sáinz & Upadyaya, 2016). Students with higher fathers' education were more likely to be assigned to the under-confident cluster than the accurately-high cluster at Grade 10 (Sáinz & Upadyaya, 2016). Students with higher mothers' or fathers' education were more likely to be assigned to the accurately-low cluster at Grade 10 than the accurately-high cluster (Sáinz & Upadyaya, 2016).

For secondary school students in Switzerland and Germany (Grade 9, age 15), students were clustered using latent-class analysis on their scores on general cognitive tests and physics content tests, their reported interest in physics, and their self-concept beliefs in physics (Seidel, 2006). Five clusters were found to be optimal to fit the data: one with generally high scores and beliefs (24% of the students); one with high cognitive scores, varying physics scores, low interest in physics, and moderate self-concept (12%); one with high cognitive scores and physics scores, moderate interest, and low self-concept (29%, potentially 'under-estimating'); one with low cognitive scores, moderate/varying physics scores, high interest, and high self-concept (16%, potentially 'over-estimating'); and one with generally low/moderate scores and responses (19%) (Seidel, 2006). The same students were considered further, highlighting that more girls were classified as under-estimating and more boys were classified as over-estimating (Jurik, Gröschner, & Seidel, 2013). The cluster with generally high scores and beliefs were observed to have higher frequency and duration of engagement and giving answers in lessons; membership of the under-estimating cluster only associated with lower frequency of engagement (and not duration of engagement or frequency of giving answers), while membership of the over-estimating cluster did not associate with students' engagement (Jurik, Gröschner, & Seidel, 2013). Additionally, the cluster with generally high scores and beliefs reported higher use of particular learning strategies (such as rephrasing content into the students' own words) and intrinsic motivation for learning, while the over-estimating students reported moderately, and while the under-estimating students reported lower (Jurik, Gröschner, & Seidel, 2014).

In summary, it remains relatively unclear, given the varying ages and samples of students considered, whether biases towards under-confidence and over-confidence occur in equal proportions of students, or only in minorities of students. The fundamental points of note were that considering clusters of students appeared to offer the potential for new insights, but appeared not to have been undertaken for students in England.

Section 3.3: Self-regulation

Motivational theories, such as social-cognitive theory applied through the expectancy-value model, can help provide interpretative and explanatory perspectives onto students' intentions and choices (Bøe & Henriksen, 2015; Eccles, 2009). Similarly, the wider perspective of self-regulation can potentially help contextualise, interpret, and/or explain the relevance of accuracy/bias in students' confidence. Indeed, the expectancy-value model, which encompasses students' confidence beliefs, itself operates within social-cognitive theory, which has been described as inherently involving self-regulation (Bandura, 1989, 2001).

For example, someone's self-efficacy and expectations have been theorised to determine, together with various other influences, what goals are chosen and what actions are taken to realise them; various beliefs, ideas, and conceptions may influence goals, actions, and the standards against which actions and results are evaluated (Bandura, 2001). Someone then monitors and evaluates their progress against their goals, which helps determine any changes in actions, motivation, and other aspects, in order to help meet the goals or even to change the goals themselves (Bandura, 2001). These various processes may occur on different levels of detail and over time, broadly within iterative cycles of 'self-regulation' (broadly covering planning, acting, monitoring, and evaluating), and may be influenced in various ways by beliefs, feelings, and affective reactions (Bandura, 1989, 2001; Butler & Winne, 1995; Zimmerman, 2000; Zimmerman & Moylan, 2009). Self-regulation, considered as a cyclical process in this way, appears conceptually similar to (biological) feedback processes (Cannon, 1929; Carver & Scheier, 1982) and general ideas of experiential learning (Kolb, 1984).

Ultimately, someone would need to accurately determine their current situation in order to monitor their progress against their goals. For example, someone would need to consider their confidence in their expected attainment, including their understanding of various topics and similar aspects, in order to determine whether further revision would be necessary in order to gain their desired examination grade. Accurate self-evaluation may then have important implications to students' studying approaches and motivation: students may study less if they believe that they already master an area, for example, which becomes problematic if this belief is inaccurate (Winne, 1995). Generally, effective self-regulation has been assumed to result in improved learning and performance, and also in accurate beliefs of ability, efficacy, or performance (Boekaerts, 1999; Pintrich, 1999).

'Self-regulated learning', considered as a specific concept in itself, has been defined as someone using deliberate cycles of planning, acting, and monitoring, specifically applied to learning activities (Butler & Winne, 1995). Essentially, the idea of self-regulation (which could potentially be implicit or explicit within various areas of life) has been formalised into an explicit strategy that students may or may not apply (or may apply to varying degrees). In accordance with theoretical assumptions, studies have highlighted the apparent benefits of self-regulated studying. Reviewed across various studies, students' self-reported self-regulated studying has associated with attainment, and at a greater average magnitude than specific studying approaches such as memorisation/rehearsal, elaboration, and organisation (Credé & Phillips, 2011). More specifically, for secondary school students in the United States, those with higher attainment reported higher adaptive self-regulation skills and interest in mathematics than those with lower attainment (Cleary & Chen, 2009). For similar students, science grades also correlated with their self-reported regulation and knowledge of their studying (Sperling, Richmond, Ramsay, & Klapp, 2012). For secondary school students in Germany, self-regulated learning has also associated with enjoyment of learning (Goetz, Hall, Frenzel, & Pekrun, 2006). For secondary school students in Korea, self-regulation, together with self-efficacy (contextualised as confidence in gaining grades) and grade goals (what grade is aimed at, what is the lowest grade that they would be satisfied in getting), directly predicted attainment across various academic subjects; students' subject interest did not directly predict attainment but instead predicted self-regulation (Lee, Lee, & Bong, 2014). The one exception was science, where the same general pattern occurred, but self-regulation did not directly predict attainment in science (Lee, Lee, & Bong, 2014).

It remains relatively unclear whether self-regulated learning, as a specific strategy, leads to accurate beliefs, whether self-regulated learning requires accurate beliefs, and/or whether other relations occur (such as elements of both cases). Given that studies have often associated higher attainment with higher accuracy in beliefs and lower biases (degrees of under-confidence/over-confidence) for secondary school students (Chen, 2003; Rinne & Mazzocco, 2014) and university students (Ackerman & Wolman, 2007; Bol, Hacker, O'Shea, & Allen, 2005), it is perhaps plausible to infer that higher accuracy may similarly associate with higher self-regulation of studying. Nevertheless, any causality would remain unclear. For example, higher attainment might facilitate higher accuracy rather than higher accuracy entailing higher attainment. Equally, 'self-regulatory' studying strategies may entail higher attainment in themselves, perhaps regardless of whether they actually reflect or require accurate beliefs (and/or a cycle of activities/processes). Additionally, the 'self-regulation' underlying the accuracy/bias of someone's beliefs, including their confidence, may differ from the (perhaps more explicit or conscious) 'self-regulation' measured in terms of study strategies.

Any association between indicators of the accuracy/bias of beliefs and potential indicators of self-regulation remains somewhat unclear (Stone, 2000). For example, it remains unclear whether students with high self-regulation of their studying strategies also have high accuracy in their beliefs, as few studies have explicitly explored the area. As one example, for undergraduate students in the United States, students' confidence accuracy (confidence compared to their task performance) did not associate with their reported knowledge of their own studying strategies and how they planned and monitored their learning (Schraw & Dennison, 1994). For similar students, performance/confidence accuracy positively correlated with self-reported knowledge of their own strengths and weaknesses, learning strategies, and when and why to apply those strategies, but only for one out of two tests (Sperling, Howard, Staley, & DuBoi, 2004). For a relatively small sample of primary school students in the Netherlands, students' expressions of self-regulation when undertaking tasks (e.g. coded to reflect planning, monitoring, reflecting on the answer, and other theoretical aspects) and their performance/confidence accuracy both positively correlated with performance but only modestly correlated with each other, and neither measure correlated with self-reported self-regulated studying (Jacobse & Harskamp, 2012).

In summary, the idea of self-regulation assumes that accurate beliefs are integral to personal well-being and functioning (Butler & Winne, 1995), hence under-confidence and over-confidence may equally be problematic. Alternately, the theorised motivational benefits of confidence (Bandura, 1997) may suggest that higher confidence, and perhaps even some degree of over-confidence, may be beneficial. Research within the area then has the potential to inform wider theoretical assumptions. The idea of self-regulation may suggest why accurate beliefs are beneficial, for example through potentially facilitating cyclical processes of learning and in working towards various goals (Bandura, 1989, 2001; Butler & Winne, 1995; Zimmerman, 2000; Zimmerman & Moylan, 2009). Specifically, without accurate beliefs it may be difficult to determine progress and to refine any studying approaches. Nevertheless, the ideas of accurate beliefs and self-regulation are distinct from the idea of an explicit and formalised process of self-regulated learning; while self-regulated learning strategies conceptually rely on accurate beliefs, any empirical association remains unclear. Self-regulated learning strategies have nevertheless broadly associated with students' attainment (Credé & Phillips, 2011). Indicators of learning or studying strategies may perhaps provide further means to help determine whether under-confident or over-confident beliefs are beneficial or detrimental.

Section 3.4: Issues inherent to considering confidence accuracy/bias

The accuracy/bias of judgments was historically considered within the specific context of probability assessment (Brier, 1950; Lichtenstein, Fischhoff, & Phillips, 1982). In order to consider its accuracy, a ‘subjective probability’ was compared against an external or ‘objective probability’, where both referred to the same event or situation; for example, the subjective probability associated with a particular weather forecast was compared to the weather on the relevant day (Brier, 1950).

Subsequent studies have explored accuracy/bias in various contexts, generally comparing someone’s predictions or retrospective evaluations of their performance (their ‘subjective’ judgments) against their actual performance (an ideally ‘objective’ situation). Various research traditions have developed, and have applied a large array of different approaches to measure performance, including non-curricular tasks, general knowledge questions, and memorisation tests of pairs of words (Hansson, Rönnlund, Juslin, & Nilsson, 2008; Koriat, 1997; Kruger & Dunning, 1999). Similarly, students’ judgments have been measured in various ways, including directly in relation to their performance or as relative comparisons or rankings against the perceived performance of others, which has sometimes entailed variable results (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008; Kruger & Dunning, 1999). Various studies have also faithfully followed the tradition of directly considering ‘probabilities’, although subsequently found that asking students to directly express or interpret probabilities can introduce ambiguity and entail variable results (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Hansson, Rönnlund, Juslin, & Nilsson, 2008; Soll & Klayman, 2004; Stankov, Lee, & Paek, 2009).

Within educational and motivational research, considering accuracy/bias has broadly developed into considering the degree of correspondence between someone’s confidence, or other beliefs about their abilities, and their attainment, or other evidence of their abilities (Pieschl, 2009; Stone, 2000). While attention has been given to applying contextually-relevant measures of each, some general methodological issues or potential limitations may be unavoidable.

Considering the degree of correspondence between someone's beliefs and their attainment assumes that a belief is inaccurate when it does not correspond to the selected indicator. This may follow from, and impose, ontological realism (where external things are as they are, independently of how they are perceived by people) and a correspondence theory of truth (where true beliefs correspond to reality) (Audi, 1998). When considering the accuracy of students' beliefs compared to their attainment, a student's justification for holding a belief (in an epistemological sense) is not necessarily explored. Justifications for holding beliefs may include correspondences with reality, inferences from other beliefs, and/or the general coherence between beliefs (Audi, 1998). Someone may feel justified in their confidence due to it cohering with their other beliefs, which may involve notions of someone's wider identity, for example; under-confidence or over-confidence may perhaps partially follow from aspiring towards or avoiding a particular identity. Alternately, confidence biases may perhaps follow from particular motivations, such as to self-enhance through maximising positive beliefs or to self-protect through minimising negative beliefs (Alicke & Sedikides, 2009; Sedikides & Alicke, 2012). Someone may also simply have low self-reflection, without any specific motivation to self-enhance or self-protect. Essentially, indicators of accuracy/bias involve specific assumptions, and do not necessarily explain why biases occur.

The correspondence between someone's confidence and their attainment may be interpreted as conceptually or theoretically representing self-reflection, self-awareness, and/or self-regulation of some kind (Pieschl, 2009; Stone, 2000). However, indicators of accuracy/bias cannot determine whether and how processes such as self-evaluation occur, and may only indirectly reflect or represent them. For example, students might form their self-concept beliefs through reflecting on their classwork or homework experiences and results rather than the specific attainment grades considered within a research study. Operationally, the accuracy/bias of someone's confidence compared to their attainment is an artificially created indicator, given particular data; different data, such as different attainment measures, may provide different results. Essentially, an indicator may not necessarily reflect any aspect of someone's thinking (Cronbach & Furby, 1970).

Conceptually and methodologically, comparing an expression of confidence and a measure of attainment requires that they both reflect the same underlying aspect (e.g. knowledge and skills in science), so that the comparison is valid. Intuitively, validity appears clear when considering confidence inherently matched to specific circumstances (e.g. test items paired with confidence-ratings such as ‘How confident are you that your answer is correct?’). Similarly, expressions of self-concept beliefs are theorised to develop from, and often empirically measured in relation to, attainment experiences (e.g. agreement with ‘I get good grades in science’) (Bong & Clark, 1999; Bong & Skaalvik, 2003). Intuitively, it appears valid to compare self-concept beliefs against attainment grades in science. Nevertheless, validity is reduced when the two measures do not exactly overlap and/or are variously influenced by other factors, such as varying interpretations of what ‘good grades’ entail.

On a general level, anything that reduces the correspondence between someone’s confidence and attainment may potentially give the appearance of under-confidence or over-confidence (Erev, Wallsten, & Budescu, 1994). For example, someone may have performed higher or lower than usual by chance or due to some features of the test items, which may give the appearance of biased confidence at the time of testing while it might appear accurate on other occasions. When averaged across numerous students, random chance alone might not necessarily invalidate the consideration of accuracy/bias, but instead might reduce reliability (i.e. it would be harder to distinguish consistent results from random variation). However, any form of systematic variation could complicate the measurement of accuracy/bias. Inherent aspects of measurement may influence someone’s responses and hence perhaps give the appearance of confidence biases (or influence their magnitude). For example, general tendencies of people to agree to items regardless of their content and/or to respond differently to positively-phrased or negatively-phrased items may be relevant, as these may influence expressions of confidence (Cronbach, 1950; Paulhus, 1991; Weems, Onwuegbuzie, & Collins, 2006; Weems, Onwuegbuzie, Schreiber, & Eggers, 2003). Such issues, however, introduce uncertainty into measurement in general, not simply measurement of accuracy/bias, and it remains somewhat unclear how prevalent particular response styles or tendencies may be.

In summary, considering confidence accuracy/bias involves comparing indicators of confidence and attainment. Validity may depend on the indicators conceptually reflecting the same area, and would be enhanced by the indicators being directly comparable or explicitly matched (e.g. comparing someone's confidence rating for a particular task against their performance for that particular task). Even when a comparison is conceptually valid, it cannot necessarily confirm that students self-evaluate in the same way, unless the expression of confidence is explicitly formed in reference to a specific indicator of attainment.

Section 3.4.1: Creating indicators of accuracy/bias

Various indicators can be calculated to quantify overall accuracy and/or bias in confidence, and/or other related aspects such as accurately determining when answers are either correct or incorrect (Lichtenstein, Fischhoff, & Phillips, 1982; Yaniv, Yates, & Smith, 1991; Yates, 1990). Many indicators have been found to be complementary, each providing specific but informative results (Boekaerts & Rozendaal, 2010).

Practically, many studies have applied multiple tasks each paired with confidence-ratings (i.e. 'How confident are you that your answer is correct?'), where the two measures can then be intuitively and directly compared through calculating a simple 'difference-score' after equalising the measures onto equivalent scales (Chen, 2003; Pajares & Graham, 1999). A difference-score broadly assumes that the equalised scales have equivalent ranges (i.e. the lowest confidence value equates to the lowest attainment value, and similarly for the highest values) and intervals (i.e. confidence increases in the same increments to attainment). This allows a symmetrical relation where over-evaluating confidence given a level of attainment is equivalent to under-attainment given a level of confidence, both in principle and in magnitude.

Calculating the simple difference between two indicators has often been assumed to reduce reliability (Cronbach & Furby, 1970). For example, the reliability of a difference-score is a function of the reliabilities of both of the two components and the correlation between them, and any correlation is likely to be imperfect (Cohen & Cohen, 1984; Willett, 1988). However, lower reliability in that situation may not necessarily reflect reduced precision or increased statistical error (Rogosa & Willett, 1983; Rogosa, Brandt, & Zimowski, 1982).

When confidence and task-scores have been both operationalised as, or can be converted into, binary measures (i.e. confident/not-confident and correct/incorrect), then numerous indicators of accuracy can be calculated from cross-tabulating the measures and considering the various cell frequencies and row/column totals: for example, ‘sensitivity’ provides a measure of when someone accurately believes themselves to be correct and ‘specificity’ provides a measure of when someone accurately believes that they are incorrect (Yule, 1912). Linking with statistical and research approaches, ‘specificity’ subtracted from one represents a Type I (false positive) error rate, and ‘sensitivity’ subtracted from one represents a Type II (false negative) error rate (Cohen, Manion, & Morrison, 2007). Such distinctions also link with theoretical assumptions that different types of confidence evaluations may be formed through different cognitive processes (Koriat, 2012). Explorations using simulated data (Schraw, Kuch, & Gutierrez, 2013) and responses from university students (Schraw, Kuch, Gutierrez, & Richmond, 2014) have highlighted that measures of sensitivity and specificity accounted for most of the variance across various other calculated binary indicators, and hence may be useful to apply.

When paired tasks and confidence-ratings have not been available, studies have compared students’ broader expressions of confidence and indicators of attainment in various ways. For example, students’ confidence has been predicted using their attainment (via linear regression), and the regression-residual has been used as an indicator of confidence accuracy/bias (Bouffard, Vezeau, Roy, & Lengelé, 2011; Gonida & Leondari, 2011; Narciss, Koerndle, & Dresel, 2011). This approach is especially useful for considering expressions of confidence (e.g. self-concept beliefs) that are not inherently contextualised against measures of attainment.

A regression-residual indicator of accuracy/bias can be interpreted as the difference between someone's expressed confidence and the predicted confidence that would be expected, given their attainment and given the association between confidence and attainment across a wider sample. Essentially, the indicator highlights 'relative' under-confidence or over-confidence, compared to others within a sample. Given the underlying algebraic calculations (i.e. regression equates someone's predicted z-score confidence with the z-score of their attainment multiplied by the correlation between the two factors seen across the sample), a regression-residual indicator of accuracy/bias can essentially become a difference-score indicator through various simplifications (Cohen & Cohen, 1984; Kolen & Brennan, 2004). However, linear (ordinary-least-squares) regression does not involve symmetrical relations (Cohen & Cohen, 1984): over-confidence given a level of attainment may not necessarily be equivalent to under-attainment given a level of confidence. Alternate regression approaches are possible to give symmetrical relations (Samuelson, 1942; Woolley, 1941), but these would make the interpretation of the regression-residuals less clear (i.e. they would involve some degree of residual confidence and residual attainment).

In summary, it remains important to consider how confidence accuracy/bias can be measured. Depending on the research design, different indicators can be calculated, which may involve different conceptual ideas of under-confidence and over-confidence that may be against specific task performance or relative to samples.

Section 4: Research aims and questions

The following sections describe the overall aims of the research, in terms of applying confidence accuracy/bias as a perspective to potentially provide new insight and value (**Section 4.1**), and the specific research questions for the study (**Section 4.2**).

Section 4.1: Research aims

An increased understanding of what influences students' intentions to study science at upper-secondary school may help wider consideration of varying or imbalanced progression (**Section 1**).

The literature review broadly highlighted that science intentions may be influenced by various factors (**Section 2**), primarily students' attitudes and beliefs towards science, including their confidence (**Section 2.1**). Recent studies in England have considered various arrays of attitudes; while students' interest in science and perceived utility of science appear to be strongly predictive of their intentions, results have nevertheless varied and it perhaps remains unclear as to which factors are the most relevant (Bates, Pollard, Usher, & Oakley, 2009; Mellors-Bourne, Connor, & Jackson, 2011; Mujtaba & Reiss, 2014; Vidal Rodeiro, 2007). Recent studies in England have also considered students' confidence expressed through their self-concept beliefs but not their self-efficacy beliefs, so it remains unclear how different expressions or conceptualisations of confidence may associate with students' intentions (DeWitt & Archer, 2015; Mujtaba & Reiss, 2014). Further insight may also be gained from considering clusters or groups of students, each with different profiles of attitudes and beliefs, and/or from considering how students' intentions may be influenced in different ways for different groups or clusters of students (**Section 2.2**). These approaches appear to offer potential value (perhaps in highlighting clusters of students who are likely or not likely to progress further within science), but have only been applied outside of England (Andersen & Chen, 2016; Chow, Eccles, & Salmela-Aro, 2012; Simpkins & Davis-Kean, 2005).

Given the importance of students' attitudes and motivational beliefs, including their confidence, on their intentions to study subjects (**Section 2** and **Section 3**), new insights may be gained through exploring whether, and how, under-confidence and over-confidence associate with students' science intentions. Accuracy/bias may be a potential barrier or facilitator, since students' confidence accuracy/bias may associate with their attitudes and beliefs (**Section 3.2**), although any particular associations and patterns of difference remain unclear. In general, the accuracy/bias of students' confidence remains under-explored in England: the majority of studies explicitly exploring confidence accuracy/bias have been undertaken in the United States (Cheema & Skultety, 2016; Chen, 2003; Chen & Zimmerman, 2007; Rinne & Mazzocco, 2014) and various countries in Europe (Dupeyrat, Escribe, Huet, & Régner, 2011; Gonida & Leondari, 2011; Rytkönen, Aunola, & Nurmi, 2007; Schwab & Hessels, 2015). Psychologically-orientated studies with undergraduate students have, again, mostly been undertaken in the United States (Ackerman & Wolman, 2007; Bol, Hacker, O'Shea, & Allen, 2005; Hacker, Bol, Horgan, & Rakow, 2000; Maki, Shields, Wheeler, & Zacchilli, 2005). Additionally, most studies have focused on academic subjects other than science, although with some exceptions (Cheema & Skultety, 2016; Seidel, 2006). Again, insight may also be gained through considering confidence accuracy/bias via empirically clustering students in addition to defining 'conceptual' groups of under-confident, accurate, and over-confident students (Rytkönen, Aunola, & Nurmi, 2007; Sáinz & Upadyaya, 2016).

Prior studies considering confidence accuracy/bias in England have been limited. For example, an early study in England considered small numbers of secondary school students in London, and compared the students' perceptions of their relative abilities compared to their peers against their relative test performance; the students were mostly accurate in their beliefs of mathematics and reading abilities, although few statistical tests were applied (Blatchford, 1997). Another study explored the accuracy of secondary school students' predicted attainment (made in Year 7/8) compared to their subsequent GCSE attainment; the students appeared mainly accurate in their expectations, although no statistical tests were applied (Attwood, Croll, Fuller, & Last, 2013). Both studies interpreted relative proportions of under-confident, accurate, and over-confident students and inferred gender differences, specifically that girls could be more under-confident (Blatchford, 1997) and that boys could be more over-confident (Attwood, Croll, Fuller, & Last, 2013). In the absence of statistical tests and given the small numbers of students considered, however, the situation remained unclear.

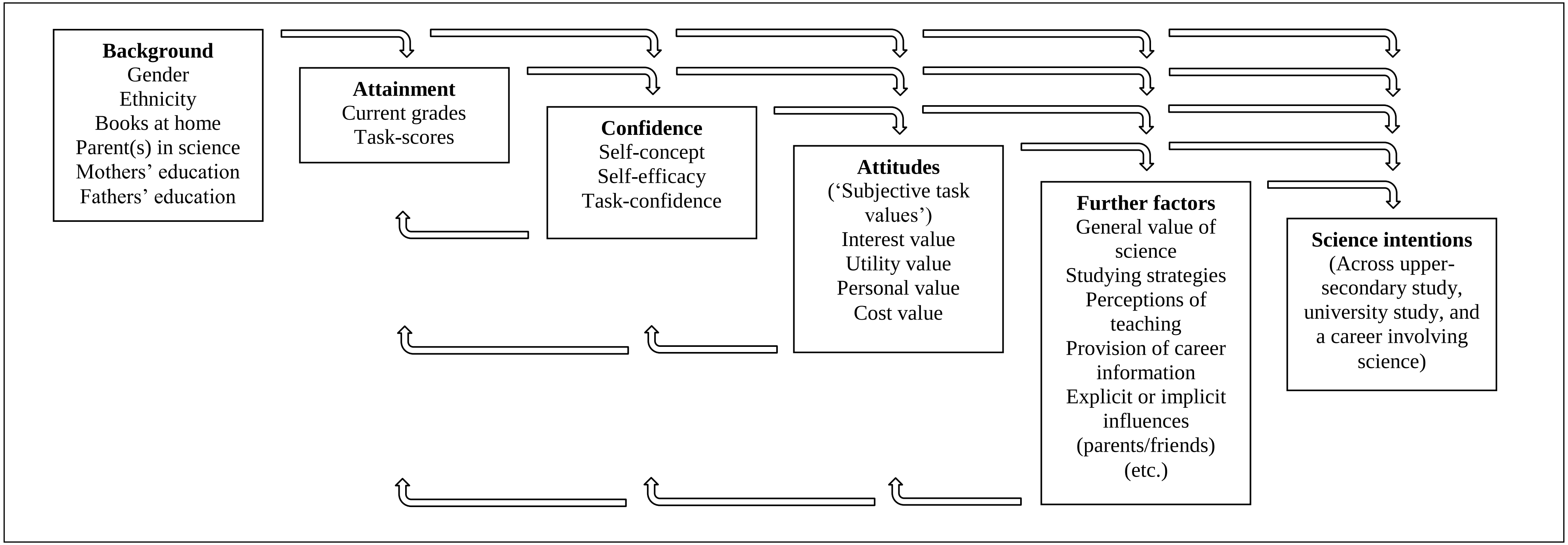
Other studies have considered the alignment of various beliefs reported by secondary school students in England, such as low or high aspirations (considered as intentions to continue into upper-secondary school), expectations (considered as a students' reported likelihood of successfully applying to university), and attainment (Khattab, 2015). Specifically, alignment between high aspirations, expectations, and attainment associated with the highest predicted probability of applying to university (Khattab, 2015). Similarly, differences between secondary school students' highest level of expected education and their career aspirations (given the researchers' assumptions about the minimum level of education required for different careers) highlighted that those with high and aligned educational and career aspirations subsequently earned higher wages (Sabate, Harris, & Staff, 2011). Such studies, however, have not explicitly considered accuracy/bias for students' confidence, and relied on researchers determining what comparisons were meaningful.

Fundamentally, the research presented in this thesis broadly aimed to provide value through offering an extended perspective on students' choices, where the implications of under-confidence and over-confidence have seldom been considered, especially for science education in England. Students' confidence was essentially considered as an 'analytical perspective', similarly to how other studies have applied students' characteristics such as gender or ethnicity/background (Archer, Halsall, & Hollingworth, 2007) and/or conceptual ideas such as identity (Archer, et al., 2010) to gain insight. However, the research broadly focused on an empirical exploration, rather than applying, integrating, or developing theory.

An illustrative conceptual model of potential influences on students' intentions, contextualised to this thesis, is provided in **Figure 1**. The model broadly links factors associated with observed differences in science intentions, specifically students' background characteristics, attainment, confidence, and attitudes such as interest and perceived utility of science (**Section 2.1** and **Section 2.2**). Following social-cognitive theory implemented through the expectancy-value model, the various factors are assumed to associate reciprocally, for example where higher attainment may lead to higher confidence beliefs, and concurrently, where higher confidence may be motivational and facilitate higher subsequent attainment (**Section 2.3** and **Section 3.1**). The model contains multiple reciprocal associations, such as between attainment and confidence, and then between attainment, confidence, and attitudes, together with any background influences. Accordingly, any initially-observed associations with intentions (such as between students' background and intentions) may reflect underlying differences in other factors (e.g. attainment, confidence, and/or attitudes) (DeWitt, Archer, & Osborne, 2014; Mujtaba & Reiss, 2014).

In addition (not illustrated for simplicity in **Figure 1**), science intentions conceptually lead to choices (which are also potentially facilitated or constrained by other factors), leading to changes in someone's environment and context, leading to changes in someone's experiences, attitudes, and beliefs, within wider cycles. Various other factors and their associations may also be relevant; for example, further factors (such as students' perceptions of their lessons) may associate directly with intentions, and/or these factors may also associate with students' confidence and/or other factors, which then associate with intentions (Hampden-Thompson & Bennett, 2013; Wang, 2012). The modelled placement of the factors is only illustrative, especially given that confirming or developing structural models was not a focus of the thesis (otherwise, such models might attempt to determine, for example, whether interest primarily leads to confidence, whether confidence primarily leads to interest, and/or whether elements of both situations occur, together with exploring various other potential combinations of associations including classroom and other experiences).

Figure 1: A conceptual model of influences on science intentions



Notes: The conceptual model broadly follows social-cognitive theory expressed via the expectancy-value model (Eccles, 2009).

Section 4.2: Research questions

Given this background and overall aims, the research was focused into three areas of enquiry.

First, it remained somewhat unclear as to which attitudes and motivational beliefs were the most relevant influences on students' intentions and choices, including different conceptualisations of confidence and indicators of confidence accuracy/bias.

Various studies have associated students' interest, perceived utility, attainment, and expressions of confidence with their intentions and choices (**Section 2.1**). Research has broadly established the motivational role of self-efficacy beliefs, in relation to numerous aspects of studying (Jiang, Song, Lee, & Bong, 2014; Skaalvik, Federici, & Klassen, 2015; Zimmerman & Schunk, 2011), although the motivational role of self-concept beliefs is plausible but perhaps less clearly evidenced (**Section 3.1**). Indicators of confidence accuracy/bias might also plausibly predict students' science intentions, inferring from research into mathematics (Sheldrake, Mujtaba, & Reiss, 2015), but it remained unclear whether this depended on the particular indicator of accuracy/bias that was calculated. Overall, students' perceived utility, interest, and self-efficacy beliefs were then hypothesised to be key influences on their science intentions.

Second, the research aimed to explore whether students with different degrees of confidence accuracy/bias exhibited different attitudes and beliefs, including whether they reported different science intentions.

Inferring from motivational theory, under-confidence could be hypothesised to be potentially limiting while over-confidence might be motivationally beneficial (Bandura, 1997); conversely, under-confidence and over-confidence may be equally problematic, and accurate beliefs might associate with higher attainment and/or effective studying (Butler & Winne, 1995). Prior studies have provided varying results (**Section 3.2**), but have generally associated over-confidence with higher interest and/or motivational tendencies to perform higher than peers (Dupeyrat, Escribe, Huet, & Régner, 2011; Gonida & Leondari, 2011; Gresham, Lane, MacMillan, Bocian, & Ward, 2000). Similarly inferring from prior research, confidence accuracy/bias may indeed associate with students' studying intentions (Martin & Debus, 1998; Sheldrake, Mujtaba, & Reiss, 2014). Inferring across these various areas, under-confidence was hypothesised to associate with lower intentions, interest, and motivational tendencies to perform higher than peers, while over-confidence and/or accuracy were hypothesised to associate with higher attitudes/beliefs.

Third, the research aimed to explore whether students with different degrees of confidence accuracy/bias considered their choices in different ways.

It was plausible to hypothesise that different students would consider their choices in different ways (**Section 2.2.3**); attainment may entail differences in how choices are made (Brown, Brown, & Bibby, 2008), hence students' confidence and accuracy/bias may similarly entail differences in how choices are made. Additionally, inferring from research into mathematics progression, interest might be more predictive of intentions for over-confident students, although patterns generally appeared unclear (Sheldrake, Mujtaba, & Reiss, 2015).

In summary, the research questions were as follows.

(1) Which attitudes and motivational beliefs (including expressions of confidence) were the most relevant influences on students' science intentions?

(2) Did students with different degrees of confidence accuracy/bias exhibit different science intentions, attitudes, and beliefs?

(3) Did students with different degrees of confidence accuracy/bias consider their science intentions in different ways?

Section 5: Research design and methods

The overall research design is described in **Section 5.1**. Essentially, the research considered two surveys, PISA 2006 and a new survey undertaken in 2014/2015. Either survey considered alone may not necessarily be ideal: PISA 2006 offered a nationally-representative sample, but could only provide an indirect indication of students' confidence accuracy/bias; the new survey provided explicit measures of students' confidence accuracy/bias, as applied in prior studies outside of England, but might have lower generalisation to other students due to limited resources when sampling and collecting data. The overall plausibility of the results would be supported if the results from the indirect measure of students' confidence accuracy/bias (in PISA 2006) cohered with the results from the direct measure of students' confidence accuracy/bias (in the new survey). With limited prior research covering confidence accuracy/bias in England, it was beneficial to generate the opportunity for contextualisation (and potential disconfirmation) within the research design itself.

The following sections then describe methodological details for PISA 2006 (**Section 5.2**) and for the 2014/2015 survey (**Section 5.3**), focusing on the samples of students and the measurement of their various attitudes and beliefs that were considered in the subsequent analysis, including how confidence accuracy/bias was measured for each survey.

As these sections highlight, many measured factors within the two surveys were conceptually comparable (aiming to measure the same underlying ideas), particularly for theorised factors such as self-concept beliefs. Nevertheless, the factors were operationalised with varying degrees of item-level comparability, given that the 2014/2015 survey aimed to be contextualised against (and to include items/dimensions from) various surveys, mainly PISA but also TIMSS: some factors (such as cost value, anxiety, and studying strategies) were not measured within PISA 2006, while other factors (such as self-efficacy beliefs) were intentionally operationalised differently in the 2014/2015 survey in order to maximise their contextual relevance to science education and progression (**Section 5.3**). In addition to the following sections that describe the measured factors for both surveys, **Appendix 1** provides a detailed reference and illustrates the item-level comparability when measuring self-concept beliefs, interest value, utility value, and personal value (potentially-relevant areas identified in **Section 2**). **Appendix 1** then also lists the items per factor for both surveys.

Section 5.1: Research design

In England, the compulsory stage of secondary school currently covers Years 7 to 11 (ages 11/12 to 15/16). During Year 9, students select various subjects to study during Years 10 and 11 at General Certificate of Secondary Education (GCSE) or equivalent level, where science is compulsory (Department for Education, 2014). Students can then undertake upper-secondary education in Years 12 and 13 (ages 16/17 to 17/18) at Advanced Level General Certificate of Education (A-Level) or equivalent level, where science subjects are optional.

Considering the prospective intentions and aspirations of students from Years 9 to 11 may increase understanding, potentially sufficient to inform practical guidance or the promotion of science. These students were also targeted to increase contextualisation against existing research that has considered students in Year 9 (DeWitt & Archer, 2015) and Year 10 (Mujtaba & Reiss, 2014). Similarly, TIMSS has focused on Year 9 (Martin & Mullis, 2013) and PISA has focused on Year 10/11 (Bradshaw, Sturman, Vappula, Ager, & Wheater, 2007; OECD, 2009a). Some uncertainty was still expected to be unavoidable: students’ prospective intentions may not necessarily reflect their subsequent choices, while students’ retrospective accounts may sometimes involve reinterpretations or rationalisations; neither approach may comprehensively reflect complex and continuous processes of decision-making (Holmegaard, Ulriksen, & Madsen, 2014). However, it was less clear how any influences of confidence accuracy/bias could have been considered in retrospect for older students already studying A-Level subjects in Year 12 and Year 13.

Students’ science intentions and attitudes can be efficiently measured across large numbers of students via surveys (DeWitt, Archer, & Osborne, 2014; Mujtaba & Reiss, 2014). Surveys can also easily include attainment tasks, paired with confidence-ratings, from which specific indicators of accuracy/bias can be calculated (Boekaerts & Rozendaal, 2010; Lichtenstein, Fischhoff, & Phillips, 1982; Yaniv, Yates, & Smith, 1991; Yates, 1990). While interviews can provide extensive detail on students' prospective intentions or retrospective choices (Holmegaard, Ulriksen, & Madsen, 2015) and what might influence their confidence (Zeldin & Pajares, 2000; Zeldin, Britner, & Pajares, 2008), they are resource intensive, which would limit the numbers of students that could be interviewed (who may not necessarily be generalizable to other students). It would also be harder to reliably associate confidence accuracy/bias with students' other attitudes and undertake predictive modelling given fewer students.

Considering existing survey data may allow responses from large numbers of students to be considered, but the data may be less suited to address particular research questions. Collecting new data is resource intensive, which may reduce the number of participating students and limit generalisation from the results, but allows the methods to be adapted to any overall aims and research questions. As a compromise, the overall research design considered two sets of student data. Fundamentally, the design aimed to address methodological limitations in one set of data with strengths in the other. The results could then be compared, and any similarities would enhance their plausibility.

First, existing data from PISA 2006 were considered, which surveyed a nationally-representative sample of students in England and considered a broad array of students' attitudes and beliefs in science, including their studying intentions (OECD, 2009a). Compared to other existing data (e.g. the Next Steps / Longitudinal Study of Young People in England, the various national life-long cohort studies), PISA studies covered wider arrays of attitudes and beliefs, were more recent, and measured students' confidence and attainment at the same time so that indicators of accuracy/bias could plausibly be calculated. Otherwise (e.g. in Next Steps), comparing indicators of confidence and attainment measured at different times could mix confidence accuracy/bias with changes over time; for example, someone might be accurate even if they reported higher confidence than might be expected, given their lower earlier attainment, if their attainment had subsequently increased. While the PISA 2006 data were older than other PISA surveys, data from PISA 2015, which also focused on science, were not available during the study, while PISA 2012 focused on mathematics and PISA 2009 focused on reading.

Second, new survey data were collected in order to measure students' self-reflective confidence accuracy/bias through paired tasks and confidence-ratings, similarly to prior research undertaken outside of England (Chen, 2003; Chen & Zimmerman, 2007). If similar patterns of results emerged across both surveys, it would support their plausibility. Essentially, with less research into confidence accuracy/bias having been undertaken in England, PISA 2006 provided a plausible baseline; however, it would be unclear whether PISA 2006 was actually measuring self-reflective accuracy/bias as considered in most prior research without comparison against other data using different methods. Conversely, the generalisation of any small new survey undertaken alone could be unclear, which could be problematic when exploring relatively new areas (e.g. whether students' intentions might be influenced in different ways depending on the confidence accuracy/bias). Additionally, given the opportunity for new data collection, the survey was designed in order to facilitate ancillary research that could provide wider contextualisation, such as what theorised influences or antecedents might associate with students' expressions of confidence and/or accuracy/bias (which is reported elsewhere, given the focus on students' science intentions within this thesis).

Little research has considered confidence accuracy/bias and students' intentions, so it was difficult to determine any necessary sample sizes. To potentially reproduce the differences in (mathematics) intentions observed in Year 10 students across under-confident, accurate, and over-confident groups (Sheldrake, Mujtaba, & Reiss, 2014), at the standard significance level of .05 (α, determining the risk of Type I errors) and with a power of .80 (power = 1 – β, where β is the standard Type II error probability of .20), a sample size of 1152 (384 per group) was expected to be needed, given power/sample size calculations (Cohen, 1992; StataCorp, 2013b). Type I errors reflect 'false positives', essentially rejecting a null hypothesis when it is actually true, while Type II errors reflect 'false negatives', essentially accepting a null hypothesis when it is actually false (Cohen, 1988, 1992). PISA 2006 surveyed 4935 students in England, while 1523 students in England (across Years 9, 10, and 11) participated in the new survey, which were then likely to be sufficient to reveal potential differences in students' science intentions across accuracy/bias groups (assuming that there were differences to be found). These numbers were also sufficient for reliable predictive modelling (Cohen & Cohen, 1984; Snijders & Bosker, 2012).

Ethical approval for the analysis of existing data and the collection and analysis of new data was given from the (UCL) Institute of Education.

Section 5.2: Methods: PISA 2006 survey

The Programme for International Student Assessment (PISA) consists of various surveys undertaken by the Organisation for Economic Co-operation and Development (OECD). The following sections describe the sample of students within PISA 2006 for England (**Section 5.2.1**), their various attitudes and beliefs that were measured by the OECD and which were considered in the subsequent analysis (**Section 5.2.2**), and how an indicator of confidence accuracy/bias was calculated (**Section 5.2.3**).

Section 5.2.1: Sampling

PISA 2006 targeted students aged 15, at the time of testing, within full-time education (OECD, 2007, 2009a). Schools were systematically sampled (with probabilities proportional to their size) within strata (with separate sampling of schools per regions and other strata), and around 35 students were then randomly-sampled within each school (OECD, 2009a).

Sample-weighting was calculated by the OECD to allow the complex sample to still reflect the wider population of students (OECD, 2009a, 2009b). Following analytical guidance from the OECD, the sample-weighting was re-scaled so that the sum of the sampling-weights was then equal to the number of considered students for England, and applied in the subsequent analysis (OECD, 2009b). The various results include sample-weighting, excepting that numbers of students are reported as un-weighted numbers for intuitive clarity.

Across England, 4935 students were surveyed (2532 girls and 2403 boys); the majority were in Year 11 (Bradshaw, Sturman, Vappula, Ager, & Wheater, 2007; OECD, 2007). As a brief contextualisation, students in England scored higher than the OECD-average in science and equal to the OECD-average in mathematics and in reading; boys in England scored higher than girls in science and mathematics, but lower than girls in reading (Bradshaw, Sturman, Vappula, Ager, & Wheater, 2007).

Section 5.2.2: Measuring students' attitudes and beliefs

Students' attitudes and beliefs cannot be directly observed, and are instead measured through various expressions or reports. Psychologically, attitudes are generally considered to be unobserved personal tendencies that are assumed to entail that someone acts in a particular way, in particular circumstances, meaning in practice that someone gives particular responses to questionnaire items (Cronbach & Meehl, 1955). Someone's questionnaire responses can be considered to reflect or to be caused by these various unobserved tendencies, which have various terms within statistical and quantitative modelling such as 'constructs' or 'factors'.

A realist perspective would assume that unobserved factors have an existence independent of measurement; existence would be necessary for factors to cause anything such as responses to questionnaire items (Bollen, 2002; Borsboom, Mellenbergh, & van Heerden, 2003; Edwards & Bagozzi, 2000). A utilitarian or pragmatic perspective would alternately consider factors to be dependent on measurement, as artificial ways to simplify and explain data or situations; unobserved factors would be pragmatically considered as empirical operationalisations of particular statistical models (Bollen, 2002; Borsboom, Mellenbergh, & van Heerden, 2003; Edwards & Bagozzi, 2000). Both perspectives may be informative. For example, students' interest in science, operationalised as a factor (i.e. aggregated responses across a set of items), embodies the idea that responses to particular items reflect wider personal tendencies; however, someone's responses may still somewhat depend on the particular items used.

Analytical approaches such as confirmatory factor analysis test whether a set of items can be considered to all contribute, in a statistical sense, to a theorised factor, and hence whether numerous items can indeed be validly simplified into one factor (Bartholomew, Steele, Moustaki, & Galbraith, 2008). The internal consistency or statistical reliability of a factor can be considered through indicators such as Cronbach's α (alpha) coefficient, which has various interpretations including the expected correlation between two random samples of items from the set of items being considered (Cronbach, 1951). Results from factor analysis and indicators of reliability nevertheless remain relative to the sample considered (e.g. results may vary across ages of students).

Within educational research, it remains difficult to logically progress from a conceptual definition of an attitude to necessary and/or sufficient aspects of measurement, practically considered as particular measurement items or questions to ask students. Various reviews of motivational attitudes and beliefs, as applied within research, have highlighted that theoretical definitions and operational measurement have sometimes varied (Eccles & Wigfield, 2002; Murphy & Alexander, 2000; Wigfield & Cambria, 2010). Different conceptual labels have been sometimes applied to the same measurement items, and/or the same conceptual label has sometimes been measured in various ways (Murphy & Alexander, 2000; Wigfield & Cambria, 2010).

Comprehensive processes of item and questionnaire development have been applied by the OECD (OECD, 2009a), and the various items within PISA 2006 broadly represent established attitudes and motivational beliefs within educational research (Wigfield & Cambria, 2010). Nevertheless, factors may be operationalised in slightly (or greatly) varying ways, and still be given the same descriptive label (Murphy & Alexander, 2000). For example, in PISA 2006, students' personal value of science was measured through agreement with items such as 'Science is very relevant to me', 'Some concepts in science help me see how I relate to other people', and 'I will use science in many ways when I am an adult' (OECD, 2009a). These may reflect a broader notion of personal value, compared to agreement with items such as ‘Science is important to me personally’, ‘Thinking scientifically is an important part of who I am’, or ‘Being able to do science helps me show other people who I am’ (Conley, 2012; Trautwein, et al., 2012; Wigfield & Eccles, 2000). Different operationalisations can be considered as providing different perspectives onto an underlying idea, ideally through some commonality of measurement items, and where results will ideally converge towards common findings (Messick, 1995).

Preliminary analysis was undertaken and the intended/theorised factors were indeed supported through confirmatory and exploratory factor analysis and indicators of reliability; essentially, the various items could validly be aggregated into the intended/theorised factors. **Table 1** provides a simple summary. (See **Appendix 1** for detailed item/factor lists.)

Table 1: PISA (England) 2006: items/factors, reliability

Factor/scale	Example item / description	Items	Reliability
Intentions	‘I would like to work in a career involving science’	4	.922
Self-concept	‘Science topics are easy for me’	6	.911
Self-efficacy (various areas)	‘Recognise the science question that underlies a newspaper report on a health issue’	8	.854
Interest (various areas)	Interest in learning about topics in physics, chemistry, biology of plants, geology, etc.	8	.847
Interest value	‘I am interested in learning about science’	5	.913
Utility value	‘What I learn in my science subjects is important for me because I need this for what I want to study later on’	5	.916
Personal value	‘Science is very relevant to me’	5	.830
General value	‘Science is valuable to society’	5	.775
Science activities	Frequency of watching programmes on science, reading science magazines, attending a science club, etc.	6	.777
School career preparation	‘The subjects I study provide me with the basic skills and knowledge for a science-related career’	4	.834
School career information	Available information regarding e.g. ‘science-related careers that are available in the job market’	4	.848
Teaching: interaction	‘Students are given opportunities to explain their ideas’	4	.772
Teaching: activities	‘Students spend time in the laboratory doing practical experiments’	4	.691
Teaching: investigations	‘Students are allowed to design their own experiments’	3	.753
Teaching: applications	‘The teacher clearly explains the relevance of science concepts to our lives’	4	.770

Notes: Reliability was measured through Cronbach’s alpha (α) coefficients.

Attitudes and beliefs

Students in PISA 2006 completed questionnaires that measured their attitudes and motivations towards science, and collected some information about their backgrounds (OECD, 2006a). Areas of the expectancy-value model of motivated behavioural choices were broadly covered by PISA 2006, including expressions of confidence and the subjective task values of interest, utility, and personal value (Eccles, 2009). However, measures of anxiety or other costs were not covered.

The OECD's theorised assignment of items to factors (OECD, 2009a) was verified through confirmatory factor analysis (via maximum-likelihood estimation, i.e. factor by factor) and through exploratory factor analysis (via principal-components analysis, i.e. considering emergent factors from all available items).

The OECD provided factor-scores calculated through item-response models, which essentially allowed response-categories per items to have varying 'difficulty', considered as the magnitude of the underlying factor required to endorse the particular agreement-scale category (de Ayala, 2009; OECD, 2009a). As a sensitivity check, preliminary analysis also calculated factor-scores as simple-averages of the relevant items (reversing the agreement-scale category scoring when necessary for consistency), which closely correlated with the OECD's factor-scores (e.g. $R = .996$, $p < .001$, for science intentions). Regardless of how the factor-scores were calculated, the various parameters in predictive modelling were sufficiently similar so that the same inferences would be made. In the final analytical models, the factor-scores provided by the OECD were used in order to increase contextualisation with prior research and/or published reports.

Students' intentions/aspirations towards science were measured (e.g. 'I would like to work in a career involving science', 'I would like to study science after secondary school'). This was considered as the main outcome of relevance for the predictive modelling.

Students' confidence in science was measured through subject-level expressions of self-concept (e.g. 'Science topics are easy for me', 'I learn science topics quickly'). Additionally, self-efficacy was measured, operationalised as someone's capacity to undertake various applied/everyday science tasks ('How easy do you think it would be for you to perform the following tasks on your own?', e.g. 'Recognise the science question that underlies a newspaper report on a health issue', 'Explain why earthquakes occur more frequently in some areas than in others').

Students' interest value (interest and enjoyment) in science was measured (e.g. 'I generally have fun when I am learning science topics', 'I am interested in learning about science'), together with area/topic-specific interest ('How much interest do you have in learning about the following science topics?', e.g. 'Topics in physics', 'Topics in chemistry'). Confirmatory and exploratory factor analysis highlighted that these measures of subject-level and area/topic-level interest indeed formed two separate factors. Utility value (e.g. 'Making an effort in my science subjects is worth it because this will help me in the work I want to do later on', 'What I learn in my science subjects is important for me because I need this for what I want to study later on') and personal value of science (e.g. 'I will use science in many ways when I am an adult', 'Science is very relevant to me') were also measured. Anxiety or other measures of cost were not measured, however, in PISA 2006.

Students' general value of science (e.g. 'Advances in science and technology usually improve people's living conditions', 'Science is valuable to society') and engagement in science-related activities were also measured (e.g. the frequency of watching programmes on science, reading science magazines, attending a science club).

Further aspects related to science careers were also measured, specifically students' perceptions of the school preparation for science careers (e.g. 'The subjects available at my school provide students with the basic skills and knowledge for a science-related career', 'The subjects I study provide me with the basic skills and knowledge for a science-related career') and information on science careers (students' reported degree of being informed about e.g. 'Science-related careers that are available in the job market', 'The steps a student needs to take if they want a science-related career').

The students' science learning context was also considered through the reported frequencies of various aspects of science teaching being applied. This specifically covered interaction (e.g. 'Students are given opportunities to explain their ideas', 'There is a class debate or discussion'), practical/hands-on activities (e.g. 'Students spend time in the laboratory doing practical experiments', 'Students are asked to draw conclusions from an experiment they have conducted'), student-led investigations (e.g. 'Students are allowed to design their own experiments', 'Students are asked to do an investigation to test out their own ideas'), and teaching that focused on models or applications of science (e.g. 'The teacher uses science to help students understand the world outside school', 'The teacher clearly explains the relevance of science concepts to our lives').

Further items/factors considering the students' awareness of environment issues were not included within the final models, however. Preliminary analysis highlighted that they had minimal to no association with students' science intentions, so these were omitted in help reduce the number of considered factors.

PISA 2006 also measured the students' own expected occupation at age 30. An indicator of whether the student expected a science-related career at age 30 (a binary indicator coded by the OECD from the students' free-text expected occupation) and the science intentions scale (e.g. agreement with 'I would like to work in a career involving science') moderately associated (R = .415, p < .001). Confirmatory factor analysis highlighted that the binary indicator had a low factor loading (.453) when modelled with the items for the intentions scale (which otherwise had a lowest factor loading of .826). Reliability also decreased (α = .922 to α = .797) when including this additional binary indicator. Preliminary analysis nevertheless highlighted (perhaps unsurprisingly) that the binary indicator strongly predicted the students' agreement-scale science intentions, even when modelling the various other attitudes and beliefs. Conceptually, both the agreement-scale and binary indicators measured the same area, although the intentions scale additionally reflected further aspects related to educational progression in science (e.g. 'I would like to study science after secondary school'). Practically, this potentially offered two different outcome measures.

The OECD's binary indicator was coded to consider various occupations such as aviation specialists (including pilots), architects, social workers, sociologists, and psychologists as science-related occupations, together with computing science, engineering, health, and natural/physical science occupations (OECD, 2009a). While explorations of supply and demand for occupations have sometimes considered wider areas such as agricultural sciences and architecture as science-related fields (Bosworth, Lyonette, Wilson, Bayliss, & Fathers, 2013), most educational research and commentary has not done so, and has generally focused on students interpreting 'science' themselves or has focused on the natural/physical sciences (DeWitt, Archer, & Osborne, 2014; Mujtaba & Reiss, 2014; Royal Society, 2014). The moderate correlation between the binary indicator and the agreement-scale science intentions might similarly follow from students interpreting the agreement-scale items such as 'I would like to work in a career involving science' to perhaps only mean careers within the natural/physical sciences.

Ultimately, considering science intentions as an agreement-scale that reflected multiple aspects of educational progression (i.e. studying science at the next educational stage(s) and also aiming for a science career) was most comparable to other research in England (DeWitt, Archer, & Osborne, 2014; Mujtaba & Reiss, 2014). Accordingly, the agreement-scale measure of science intentions was used as the outcome for the predictive models. Additionally, analytical constraints were such that sample-weighting and multi-level modelling features were supported in linear predictive modelling but were not fully supported in (logistic) predictive modelling of a binary outcome.

Students' background

Various aspects of students' background and context were also measured in PISA 2006. These included home possessions, including indicators of wealth, cultural possessions, home educational resources, and the number of books at home (OECD, 2009a). Preliminary analysis highlighted that the scale of home possessions produced broadly similar parameters to the item measuring the number of books at home considered alone. In order to increase general comparability with the new data collection, where it was not feasible to ask students an extensive array of questions regarding their possessions, the indicator of the number of books at home was used in the final models.

PISA 2006 also measured the students' parents' highest occupational level (coded by the OECD from the students' open-ended responses) and highest educational level (OECD, 2009a). The occupational level was considered by the OECD through the international socio-economic index of occupational status (ISEI), a continuous scale that reflected income and educational differences across occupations (Ganzeboom, De Graaf, & Treiman, 1992). Preliminary analysis highlighted that the parents' highest educational levels and occupational levels had no predictive association with students' science intentions when controlling for the students' attitudes and beliefs, except for the fathers' (or male guardians') highest educational level within some models. In order to simplify modelling, the parents' highest educational levels (and not the parents' occupational levels) were retained in the final models. This also increased model comparability with the new data collection, where it was not feasible to ask students about their parental occupations and classify the results (and this also enhanced potential comparability with other studies such as TIMSS that have only considered parental education and not occupation).

An indicator of whether the students' mother or father worked in a science-related career (yes or no/undetermined, as coded by the OECD) (OECD, 2009a) was also included in the final models. Parents working within science might reflect implicit dispositions or attitudes to science that may be promoted within families, which may then influence students' own aspirations towards science (Archer, Dawson, DeWitt, Seakins, & Wong, 2015).

Task-scores

Students in PISA 2006 completed test booklets containing various sets of applied tasks (not necessarily curricula-based), covering science, mathematics, and reading items. Not every questionnaire booklet included every task item. Students could receive one of thirteen possible task booklets, each containing different sets of items; some science items were included in every booklet, but no mathematics items were present in three booklets and no reading items were present in six booklets (OECD, 2009a). For science, the average number of items per booklet was 32 (standard deviation of 13, minimum of 15, maximum of 60, all rounded to the nearer integer; the average reliability across all booklets was $\alpha = .872$).

Students' task-scores were treated by the OECD as 'missing responses' to be analytically inferred from the tasks that were assigned, given that ideally all students would have completed every possible task. Operationally, the OECD calculated students' task-scores through item-response models (i.e. estimating scores given the particular responses to the test items), while using the students' other responses (i.e. their reported background characteristics and attitudes) as additional information to help infer 'missing responses' across all students through multiple-imputation (OECD, 2009a). Estimates of task-scores were then provided as 'plausible-values', five random-selections from each student's estimated distribution of scores (OECD, 2009a).

The plausible-values were broadly designed to optimise performance estimates for wider populations of students, rather than individual students (OECD, 2009b). Analysis would ideally be repeated and results would be combined across all five plausible-values (OECD, 2009b; Rubin, 1987). Using one single plausible-value alone would nevertheless give statistically-unbiased estimates, and remains an acceptable approach, but would include some unknown degree of additional (theoretically-random) imputation variance (OECD, 2009b). Accordingly, studies of PISA have variously used one single plausible-value (Cheema & Skultety, 2016) or combined all plausible-values through multiple-imputation (Nagengast & Marsh, 2011).

On average for students in England, the OECD plausible-values closely but imperfectly associated, with a mean of $R = .928$, $p < .001$, across the possible correlations between the five different plausible-values, which highlighted the imputation variance. This could potentially complicate considering accuracy/bias when comparing students' confidence against different plausible-values and when defining groups: someone might be differently classified as under-confident, accurate, and/or over-confident when using different plausible-values. However, approaches to combine estimates across multiple plausible-values assume that group sizes and compositions are constant across the various plausible-values (Rubin, 1987). Preliminary analysis therefore explored the implications of using one plausible-value and using all plausible-values (analytically aggregated through multiple-imputation). The parameters in predictive modelling were similar, such that conclusions and inferences would remain unchanged either way. Given that latent-profile analysis did not support multiple-imputation across multiple plausible-values, only the first plausible-value was used for consistency in the final analytical models.

As a further sensitivity check, preliminary analysis also considered single estimates of task-scores that were calculated as averages of the administered items (i.e. the mean proportion correct) and calculated through simple (single-parameter logistic) item-response models without attempting to infer/impute across the rotated design (i.e. calculating models for each of the thirteen booklets) (de Ayala, 2009; Rabe-Hesketh, Skrondal, & Pickles, 2004; Zheng & Rabe-Hesketh, 2007). These calculated item-response and simple-average task-scores closely correlated ($R = .987$, $p < .001$), and also correlated with the plausible-values (respectively $R = .920$, $p < .001$, and $R = .918$, $p < .001$, calculated through multiple-imputation across the plausible-values). Regardless of the different methods used to calculate the task-scores, the parameters in predictive modelling were broadly similar so that fundamental conclusions or inferences would not change. However, any calculated task-score could not match the complexity of the item-response models applied by the OECD (i.e. modelling both binary and partial-credit tasks) (OECD, 2009a). As above, the first plausible-value from the OECD was therefore used in the final analysis.

Section 5.2.3: Measuring students' confidence biases

The design of PISA 2006 provided measures of confidence (i.e. self-concept and self-efficacy) and one measure of attainment (i.e. task-score). Preliminary analysis highlighted that the students' science self-concept beliefs had a higher predictive association with their science intentions than their self-efficacy beliefs for the students in England, which cohered with prior research considering students in Germany (Jansen, Scherer, &

Schroeders, 2015). Accordingly, the accuracy/bias of the students' self-concept beliefs formed the focus of the subsequent analysis (i.e. given that self-concept beliefs appeared to have greater contextual relevance). While the task-scores covered applied areas and may not perfectly reflect the students' classroom (curricular-based) attainment, alternate attainment measures were not available; nevertheless, when PISA cohorts have been considered longitudinally in various countries, PISA task-scores have positively associated with students' subsequent examination attainment with small to moderate magnitudes (Fischbach, Keller, Preckel, & Brunner, 2013; Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014).

In PISA 2006 for England, students' self-concept beliefs correlated to some extent with their task-scores (R = .331, p < .001, for the first plausible-value alone, or R = .330, p < .001, across all plausible-values via multiple-imputation). This relatively modest association could reflect various situations. First, students may form their self-concept beliefs in reference to diverse measures of attainment that are only approximated by the (applied and non-curricular) OECD tasks; essentially, the OECD tasks may not be ideal for determining self-concept accuracy/bias. Second, students' beliefs may be influenced by factors other than attainment (**Section 3.1**), and imperfect associations are then perhaps unavoidable. Third, some students may be variously under-confident or over-confident in their self-concept beliefs, when compared to their task-scores (**Section 3.2**; although it remains unclear whether the same accuracy/bias would be apparent when compared to other indicators of attainment). All three situations may occur to some extent (plus other potential conceptual/theoretical issues may be relevant, as discussed in **Section 3.4**). Accordingly, and unavoidably, there is some uncertainty and/or imprecision in any exploration of confidence accuracy/bias, although this does not necessarily make the process invalid or without potential benefit.

Considering the accuracy/bias of students' self-concept beliefs against their task-scores required a method to associate the two measures (**Section 3.4**): for example, task-scores could simply be subtracted from confidence in a simple difference-score, assuming that the two indicators were measured on equivalent scales. However, students' expressions of self-concept beliefs (e.g. agreement with 'Science topics are easy for me') were not inherently contextualised or referenced against task-scores. For example, science being 'easy' does not necessarily entail attaining at a particular task-score; different students may also have different interpretations of what 'ease' entails.

In order to directly compare self-concept/self-efficacy and task-scores, prior research has sometimes standardised these indicators (i.e. through *z*-score transformations); standardised indicators then intuitively appear to be on the same scale (i.e. standard deviations above or below the mean) and directly comparable. For example, one prior study of PISA calculated an accuracy/bias index as the students' expressed self-concept beliefs minus the students' task-score, minus the country-mean task-score (multiple countries were considered); the components were standardised via *z*-score transformations across all countries (Chiu & Klassen, 2010). Similarly, another study calculated an accuracy/bias index as the students' expressed self-efficacy minus the students' task-score, where both components were standardised via *z*-score transformations across the country being considered (Cheema & Skultety, 2016). These indicators can be broadly interpreted to reflect relative accuracy/bias, given the sample. For example, accuracy would entail confidence beliefs and attainment both being similarly above or below the sample mean, in standard deviation units.

However, 'relative accuracy' may not necessarily entail confidence increasing by one standard deviation when attainment increases by one standard deviation. Across a sample, the particular association between confidence and attainment can be revealed through, for example, linear regression (i.e. which also accounts for the correlation between confidence and attainment). 'Relative accuracy' would then entail someone's expressed confidence corresponding to the prediction from the regression equation, and under-confidence or over-confidence would be revealed by the regression-residual. For example, under-confidence would be revealed through lower reported confidence than predicted confidence and a negative regression-residual. Prior research has considered accuracy/bias through regression-residual indicators in this way, although not using PISA samples (Bouffard, Vezeau, Roy, & Lengelé, 2011; Gonida & Leondari, 2011; Narciss, Koerndle, & Dresel, 2011).

Accordingly, a regression-residual indicator of accuracy/bias was used to compare the students' self-concept beliefs and task-scores (for consistency across all the final analysis, this only considered the first plausible-value). This can be interpreted as the difference between someone's expressed self-concept and the predicted self-concept that would be expected, given their task-scores and given the association between self-concept and task-scores across the sample. A regression-residual indicator is essentially equivalent to a difference between *z*-scores, but also accounts for the correlation between the two factors (**Section 3.4**).

Section 5.3: Methods: 2014/2015 survey

In order to comprehensively address the research aims (**Section 4**) a new survey was developed and applied in 2014/2015. In PISA 2006, confidence accuracy/bias could only be considered relatively imprecisely through comparing non-matching indicators of confidence and attainment. The new survey was applied in order to determine whether similar results would be observed when considering explicit indicators of students' self-reflective confidence accuracy/bias, measured via matched tasks and confidence-ratings as applied in various prior studies (Chen, 2003; Pajares & Graham, 1999).

The following sections cover the sampling in **Section 5.3.1** and development of the questionnaire in **Section 5.3.2**. The various science attitudes and beliefs that were measured and used within the analysis are detailed in **Section 5.3.3**, and the measures of confidence accuracy/bias are similarly detailed in **Section 5.3.4**.

Section 5.3.1: Sampling

Surveying students usually requires schools to be sampled and approached, since students cannot easily be directly invited to participate in research. Schools may be classified into various types, depending on observable features such as their admissions policies; schools may also have varying contextual features, such as their teaching approaches, which may not easily be observable or described within national records.

Stratified sampling can help ensure that sufficient numbers of participants are gathered from specific types of schools; however, this requires determining what strata are relevant to consider and then requires larger samples to adequately cover all strata (Cohen, Manion, & Morrison, 2007). The overall research design focused on students' attitudes and beliefs, rather than the influence of types of school or other aspects that could easily form strata. Additionally, schools needed to be sampled and invited in stages (e.g. for piloting and then at subsequent stages), with limited resources and with no way to guarantee participation within strata or otherwise. Accordingly, it was operationally more feasible to randomly-sample schools, although this potentially limited generalisation, depending on participation.

National attainment records for Key Stage 4 (GCSE and equivalent qualifications) for secondary schools in England as of 2012/2013 (the latest available during 2014) formed the sampling frame, and data collection occurred during the 2014/2015 academic year. Subsequent national attainment records from 2013/2014 and 2014/2015 were considered for further contextualisation when they became available (Department of Education, 2016).

The sampling frame excluded schools exclusively for those with special educational needs, who might find a questionnaire inaccessible and/or might have limited generalisation to other students. The sampling frame therefore considered 4125 schools out of 5238 secondary schools across England as of 2012/2013. Schools were then randomly selected and invited to participate, regardless of school type, admissions policies, and other school features. Given limited resources, invitations were sent electronically (and not via post); if there was no feasible means to contact a school, another was randomly selected instead. Schools were invited in stages (i.e. inviting 10-20 schools and allowing a few weeks for responses before inviting more) until resources were exhausted, for example on questionnaire printing and courier costs. This process was necessary to avoid potentially over-committing resources (e.g. if 100 schools were invited at once, and all chose to participate, then there would have been insufficient resources to provide printed questionnaires for them all).

In total, 314 schools were invited and 12 schools participated for science. Schools participated at their convenience (but were reminded and prompted to facilitate a decision either way). The invitation explained the research and potential benefits of gaining more knowledge (e.g. highlighting that understanding students' confidence may have implications on guidance for subject choices or careers). Schools were also offered anonymous summaries of their students' responses (e.g. school-level averages per questionnaire item), which were subsequently provided.

From the twelve participating schools, seven were mixed-admissions comprehensive schools (admitting boys and girls, and not selecting students based on their attainment). Mixed-admissions comprehensive schools formed the majority of all secondary schools within England: 67% as of 2012/2013 and 68% as of 2014/2015 (Department of Education, 2016). Selective schools (only admitting students based on their attainment) and boys-only and girls-only schools were also represented in the sample, but with minimal numbers. Further details of the sampled schools are covered in **Appendix 2**.

The twelve schools covered a range of examination attainment evidenced by prior cohorts, although on average, as of 2012/2013, 65% of their students were reported to have achieved five or more A*-C grades (including in both English and mathematics) at GCSE level compared to averages of 64% across the invited schools, 61% across the sampling frame, and 51% across all schools in England (i.e. including schools for special educational needs). The equivalent averages, as of 2014/2015, were 62% for the sample, 59% across the invited schools, 54% across the sampling frame, and 47% across all national schools. Due to continual changes of school status (e.g. schools closing, opening, and/or changing status and hence identification codes), not all schools on the original 2012/2013 sampling frame could be matched against subsequent records. The contextualisation unavoidably involved slightly different numbers of schools for each year, and also involved changes in attainment due to different cohorts sitting the examinations.

Any sampling approach may potentially encounter different schools being more or less likely to participate (Cohen, Manion, & Morrison, 2007). Considering the 2012/2013 data, there appeared to be no clear participation bias in terms of attainment, although with only twelve schools participating any inferences cannot be conclusive. However, the invitation process appeared to (inadvertently) select slightly higher attaining schools, perhaps following from re-sampling inaccessible schools. In practice, 'random' selection may not be completely possible with limited resources.

Teachers were given freedom to disseminate the questionnaires to their students to suit their context (e.g. during free periods or during lesson time). Teachers were not supervised during the process. The introduction to the questionnaire explained the purpose of the study to the students, that participation was voluntary and that any particular items could be left blank, that the data would be kept confidential and individuals would not be identified, and that further information about the research or data could be requested.

All students within Years 9, 10, and 11 were invited to be surveyed, in order to explore confidence accuracy/bias while science was still a compulsory subject. Considering multiple ages of students was intended to allow the results to be contextualised against a wide range of existing national and international research (**Section 5.1**). Some schools also offered for some younger students to participate (i.e. in Years 7 and 8); some data were then collected, although the numbers were relatively low and were not considered in the final analysis. Potential student-level participation biases could not be considered, given no wider information (e.g. on all students per schools) and potentially different practices being applied within schools depending on the teachers (e.g. where the questionnaires may have been administered for some classes/forms, in some science lessons, or via other arrangements).

In total, 685 students in Year 9, 489 in Year 10, and 349 in Year 11 were surveyed (1523 total). Given limited resources and participation, fewer than expected students were surveyed, especially in Year 11 (where teachers/schools may have wanted to focus on examination preparation rather than on completing surveys). This introduced methodological complications, such as whether to analyse separately by academic year or across all the surveyed students, as described later within the analytical approaches (**Section 6.4**).

Section 5.3.2: Questionnaire development

The questionnaire for the 2014/2015 survey aimed to reveal whether students were under-confident, accurate, or over-confident in their confidence on various science tasks, and to measure students' attitudes and motivational beliefs that might associate with their confidence accuracy/bias and/or with their science intentions. (A questionnaire is reproduced in **Appendix 3**.) The questionnaire was conceived within social-cognitive theory, applied through the expectancy-value model of motivated behavioural choices (Bandura, 1997; Eccles, 2009), and also aimed to measure additional attitudes and/or motivational beliefs of potential relevance (Eccles & Wigfield, 2002; Wigfield & Cambria, 2010).

The questionnaire considered 'science' holistically, in accordance with the National Curriculum, for comparability with prior national and international studies, and for brevity and practicality (Department for Education, 2013, 2014; DeWitt & Archer, 2015; Martin & Mullis, 2013; OECD, 2009a). Repeating items/factors for biology, chemistry, physics, and any other subjects, would either entail an unfeasibly long questionnaire, or limit the overall amount of covered content. Nevertheless, students may have varying attitudes and beliefs across biology, chemistry, and physics (Jansen, Schroeders, & Lüdtke, 2014), and across other areas such as the nature of science and practical skills (Hardy, 2014). Fundamentally, the contextual relevance was therefore unavoidably reduced through considering 'science' rather than separate subjects.

The measurement items were developed to be broadly comparable with a range of existing research (Martin & Mullis, 2013; OECD, 2009a, 2014). Ensuring comparability by drawing on existing items and conceptualisations enhanced factor/construct validity (i.e. the various items/factors were more likely to reflect prior theoretical and operational understandings), and (theoretically) content validity (i.e. the items were more likely to represent what they intended to measure) (Messick, 1995).

Nevertheless, prior items and/or conceptualisations were still closely considered.

The initial pilot questionnaires were provided as electronic and paper formats. The majority of the subsequent questionnaires were completed by students on paper within school, as teachers highlighted that this was easier for them to administer. All paper responses were then manually recorded as electronic data. The pilot questionnaires included rating-scales and free-text prompts so that students could provide feedback. Anecdotally, the feedback could vary from students enjoying the questionnaire and recognising the potential benefit of research to education, to students highlighting that the questionnaire was uninteresting and too long. Perhaps surprisingly, the tasks were relatively well-received.

Questionnaire piloting involved 165 students across Years 9, 10, and 11, which clarified operational matters such as using paper questionnaires rather than an online format for the subsequent work, and helped to suggest initial refinements in scope and items, although with some uncertainty due to the limited numbers of students. Offering incentives to students was also explored initially in piloting (i.e. students could optionally enter their name to be randomly-selected to receive a gift voucher), but appeared to produce no substantial increase in returns; vouchers were then subsequently disseminated to the relevant students via their teachers, and all data were made anonymous if names had been entered. Incentives were not offered subsequently.

Various new items were explored through questionnaire piloting, and particular items and areas of exploration were accordingly refined and/or eliminated. For example, students may potentially respond differently to positively-phrased and negatively-phrased items, such as being more likely to agree with positively-phrased items than to disagree with negatively-phrased items (Benson & Hocevar, 1985; Lindwall, et al., 2012; Weems, Onwuegbuzie, & Collins, 2006; Weems, Onwuegbuzie, Schreiber, & Eggers, 2003). Any methodological effects (or other response tendencies) might influence students' expressions of confidence, and hence the apparent accuracy/bias of their confidence. Accordingly, the pilot questionnaire included items such as agreement or disagreement with 'I usually do well in science' and 'I am bad at science', 'Science is harder for me than for many of my classmates' and 'Science is easier for me than for other students', and various other items. However, students' free-text feedback highlighted that they easily noticed the repetition (and thought that it was tiresome, redundant, and unnecessary), and students' responses were essentially replicated but mirrored across the equivalent positively-phrased and negatively-phrased items (i.e. there appeared to be little to no difference/bias).

Essentially, it was unfeasible to continue to apply additional negatively-phrased items in this way, and eliminating these items helped reduce the questionnaire length. Historically, applying a balance between positively-phrased and negatively-phrased items was suggested in order to reduce potential response tendencies such as acquiescence (Likert, 1932). Subsequently, however, recommendations have been to simply phrase items in order to reduce any potential ambiguity (Cronbach, 1950; Haladyna & Rodriguez, 2013). Focusing on positively-phrased items may also ensure that items are generally easier to understand and increase factor reliability (Barnette, 2000; Marsh, 1986a; Sliter & Zickar, 2014). Some particular negatively-phrased items were still retained for direct comparability with prior studies, however (**Section 5.3.3**).

After the initial pilot questionnaires, some teachers still highlighted that the questionnaire was relatively long, and changes were sometimes requested (such as measuring grades through free-text boxes rather than tick-boxes to encompass National Curriculum levels, as described in **Section 5.3.3**). Accordingly, the length of the questionnaire was successively reduced where possible through further iterations, while core items were retained, in order to make the questionnaire more accessible and feasible for teachers to administer within shorter lessons. This involved attempting to refine the questionnaire using limited data (e.g. the paper questionnaires took time to be entered as electronic data). Most notably, the last version of the questionnaire (only used by 18% of the students) involved the removal of two entire tasks, and two factors (cost value and personal value) were reduced from three to two items, in order to substantially reduce the length (items/factors are detailed in **Section 5.3.3**).

Initial analysis applied a cautious approach, and formed factors only from the common/core items, retained across all questionnaire versions including the pilot versions. For the majority of students, more information was available in addition to the core items. For example, across Years 9-11, 1016 students responded to ‘Science is important to me personally’, 1298 to ‘Thinking scientifically is an important part of who I am’, and 1296 to ‘Being able to do science helps me show other people who I am’. While one item was removed in the last questionnaire version (‘Science is important to me personally’), the students’ personal value of science could still be calculated across three items for most students (but across only two items for some). Factors and task-scores were accordingly calculated given the questionnaire version to maximise the information considered; for example, average task-scores (i.e. the proportion correct) were calculated out of eight or ten items, depending on the questionnaire version (**Section 5.3.3**). Preliminary analysis gave similar results either way (e.g. with personal value calculated with two or three items), so the factors were created as above to maximise the available data.

Considering the final data from the 1523 surveyed students (685 students in Year 9, 489 in Year 10, and 349 in Year 11, across all questionnaire versions), the assignment of items to wider factors was confirmed and/or refined through confirmatory factor analysis (via maximum-likelihood estimation, i.e. factor by factor) and through exploratory factor analysis (via principal-components analysis, i.e. considering emergent factors from all the available items). The process balanced maximising the number of items per factor with ensuring that the items were conceptually coherent/valid and empirically associated with the underlying theorised factor or idea. For example, the factor measuring students’ perceptions of their teacher and learning context potentially included the item ‘I think of things not related to the science lesson’, as applied in prior research (Martin & Mullis, 2013). For students across Years 9-11, this item had a low association with the theorised underlying factor (a factor loading of -.010, while the other items had loadings between .672 and .817) and entailed that two factors were identified rather than the single theorised factor. Removing the item ensured that the remaining items formed only one factor, and improved the overall reliability (α = .864 to α = .904). Essentially, the item ‘I think of things not related to the science lesson’ may measure some form of distraction or disengagement, which conceptually differs from the other items (e.g. ‘My Science teacher is easy to understand’, ‘My science teacher gives me interesting things to do’, ‘My Science teacher tells me what I need to do to become better in science’). However, in order to maintain broad comparability with the factor as applied in prior studies, further potential division of the items was not applied (e.g. separating general experiences/reactions from experiences of formative feedback/guidance).

Section 5.3.3: Measuring students’ attitudes and beliefs

Factors are often developed through conceptualising or defining an area to be explored, undertaking a review of the existing literature and measurement items, selecting specific items aided by statistical analysis, and considering how the resulting factor associates with existing measures or outcomes (Clark & Watson, 1995; Simms, 2008). Ideally, the various items will be comprehensively representative of the factor (Messick, 1995). Within educational and motivational research, however, it remains difficult to conclusively determine how factors should be defined; even given a particular definition, it remains difficult to then (logically) derive necessary and/or sufficient measurement items. Instead, the development of measurement items and factors has generally balanced theoretical and empirical considerations (Messick, 1995).

Some of these issues were reduced by considering existing ideas, concepts, or factors, such as ‘self-concept’ and ‘utility value’ rather than defining and developing measurement items for entirely new concepts. The questionnaire development then involved reviewing and aggregating questionnaire items from various sources (e.g. prior instruments, studies, etc.) for specific areas/factors (e.g. self-concept, interest, utility, etc.), identifying similar items and removing duplicates, and considering and/or adjusting the phrasing of items to ensure their contextual relevance and ease of comprehension. New items were developed where the underlying idea unavoidably required more contextualisation (e.g. science intentions via ‘I intend to study science at A-Level’ or self-efficacy via ‘What grade do you think you will be able to get at GCSE (or equivalent) science’). Once data were collected, various statistical analyses then considered and helped refine the links between items and factors (e.g. correlations between items, exploratory factor analysis across all potential items, confirmatory factor analysis across theorised items). Ultimately, single-factor structures (via confirmatory factor analysis) and acceptable indicators of reliability (Cronbach's α coefficients) were confirmed, for students considered per academic year and across multiple years.

The questionnaire items used agreement scales with categories of (1) 'strongly disagree', (2) 'disagree', (3) 'slightly disagree', (4) 'slightly agree', (5) 'agree', and (6) 'strongly agree'. Depending on the item phrasing, categories were reverse-scored when necessary so that high values (e.g. 6) consistently indicated a positive experience or belief (e.g. doing well, being interested, the absence of anxiety, etc.). Preliminary analysis explored calculating factor-scores as averages of the relevant items (i.e. 'observed' scores) and as predictions (i.e. estimated 'unobserved' or 'latent' scores) from confirmatory factor analysis (Bartholomew, Steele, Moustaki, & Galbraith, 2008) and one-parameter-logistic item-response models (Rabe-Hesketh, Skrondal, & Pickles, 2004; Zheng & Rabe-Hesketh, 2007). These various factor-scores appeared to closely associate. For example, across Years 9-11, and without inferring any missing responses, a simple-average factor-score for students' self-concept beliefs highly correlated with predicted factor-scores from confirmatory factor analysis (R = .999, p < .001) and from one-parameter-logistic item-response models (R = .986, p < .001). Additionally, preliminary analysis highlighted that similar conclusions could be inferred, such as when predicting science intentions, regardless of how the factor-scores were calculated.

Estimating factor-scores through predictive models entailed that the various model parameters (and hence the resulting factor-scores) were dependent on the particular sample of students, which introduced further potential decisions (e.g. whether to estimate factor-scores per year or across multiple years). Parameters might vary across academic years (or other groups) due to differences in how items were interpreted or contributed to factors, or due to variability from low sample sizes. While item-response models have been successfully applied by the OECD in PISA studies, for example, the relevant sample sizes have been large and the processes have only needed to consider one age group (OECD, 2009a). Additionally, various forms of item-response or other models can be applied (de Ayala, 2009), requiring further layers of assumptions or justifications into applying one particular model over another.

The final analysis used factor-scores calculated as averages of the relevant items, in order to avoid such issues. Essentially, each item was then implicitly assumed to have equal relevance or contribution to the wider factor; confirmatory factor analysis indeed highlighted that the various items within the various factors had acceptable and relatively similar factor loadings (e.g. items with low loadings were removed during the development/refinement process). This also insured that all the various item/factor-scores, including those only covered by single-items, remained on the same metric: all were observed scores, rather than some being observed scores (such as the single-items) and others being predicted scores from confirmatory factor analysis or item-response models. Fundamentally, this helped reduce the layers of methodological complexity and assumptions, especially given that research in science education has similarly focused on students' observed responses/scores (DeWitt & Archer, 2015; Mujtaba & Reiss, 2014).

The various factors are summarised in **Table 2** and further elaborated in the following sections, together with the relevant single-item indicators. **Table 3** provides an example of item-level comparability across surveys for self-concept beliefs. (See **Appendix 1** for further details of item-level comparability, and also for detailed item/factor lists for the 2014/2015 survey.)

Table 2: England 2014/2015 (Years 9, 10, 11): items/factors, reliability

Factor/scale	Example item	Items	Reliability: Year 9	Year 10	Year 11	Years 9-11
Intentions	'I intend to study science at A-Level'	3	.882	.878	.880	.882
Self-concept	'I usually do well in science'	5	.894	.898	.895	.896
Self-efficacy	'What grade do you think you will be able to get at GCSE (or equivalent) science'	2	.850	.841	.804	.835
Interest value	'I am interested in the things I learn in science'	7	.942	.923	.937	.936
Utility value	'Science is an important subject for me because I need it for what I want to study later on'	11	.948	.941	.949	.948
Personal value	'Thinking scientifically is an important part of who I am'	3	.904	.876	.883	.892
Cost value	'I have to invest a lot of time to get good grades in science'	4	.873	.861	.853	.867
Perceived control	'If I put in enough effort I can succeed in science'	5	.817	.885	.867	.856
Perceived control (exams)	'I do badly in science whether or not I study for my exams'	4	.863	.876	.842	.863
Study strategy: self-regulation	'When I study for my science class, I set goals for myself in order to direct my activities in each study period'	12	.833	.810	.840	.829
Study strategy: control	'When I study science, I start by working out exactly what I need to learn'	4	.864	.805	.860	.848
Study strategy: memorisation	'In order to remember the method for solving a science problem, I go through examples again and again'	4	.784	.663	.781	.759
Study strategy: elaboration	'When I study science, I try to relate the work to things I have learnt in other subjects'	4	.837	.813	.833	.831
Anxiety	'Science makes me confused and nervous'	5	.905	.906	.902	.905
Social persuasions (praise)	'My science teacher tells me I am good at science'	3	.808	.771	.811	.797
Subjective norms (friends)	'Most of my friends do well in science'	3	.625	.658	.656	.645
Subjective norms (parents)	'My parents believe it's important for me to study science'	3	.828	.814	.779	.820
Teacher perceptions	'My science teacher gives me interesting things to do'	8	.908	.897	.905	.904
Careers/events	'My science teacher tells me about careers and jobs in science'	3	.710	.674	.764	.707

Notes: Reliability was measured through Cronbach's alpha (α) coefficients. Items/factors were consistently score/coded so that higher values reflected positive attitudes/beliefs (e.g. higher interest, the absence of anxiety, disagreement with 'I do badly in science whether or not I study for my exams', etc.). For some students, personal value and cost were only measured through two items.

Table 3: England 2014/2015 (Years 9, 10, 11): item/factor development and survey comparability example (measuring self-concept beliefs)

		Example source/reference items		
Dimension/theme	2014/2015 survey items	PISA 2006	TIMSS 2011	PISA 2012
Self-evaluation (doing well, being good, ease)	'I usually do well in science' 'I have always been good at science'	'Science topics are easy for me' 'I can easily understand new ideas in science' 'When I am being taught science, I can understand the concepts very well'	'I usually do well in science' 'Science is not one of my strengths'	'I am just not good at science'
Mastery experiences (and their interpretation/evaluation)	'I get good grades in science'	'I can usually give good answers to test questions on science topics'	(No included items)	'I get good grades in science'
Mastery experiences of difficult work	'I understand even the most difficult science work'	'Learning advanced science topics would be easy for me'	'I am good at working out difficult science problems'	'In my science class, I understand even the most difficult work'
Learning quickly	'I learn things quickly in science'	'I learn science topics quickly'	'I learn things quickly in science'	'I learn science quickly'

Notes: PISA 2012 measured students' attitudes and beliefs related to mathematics; the listed items have been re-phrased to science for illustration. The distinction between the dimensions of 'self-evaluation' and 'mastery experiences' remains flexible (i.e. even items plausibly measuring mastery experiences involve some degree of evaluation, such as what 'good grades' would entail).

Science intentions

Students' intentions towards science were measured across upper-secondary study (A-Level or equivalent), university study, and a career involving science (i.e. 'I intend to study science at A-Level', 'I intend to study science at university', 'I am planning on pursuing a career that involves a lot of science'). The average across these items then reflected the students' aspirations to persist within science across all these stages. As with PISA 2006, considering science intentions across these multiple aspects of educational and career progression was most comparable to intentions/aspirations as considered in prior research in England (DeWitt, Archer, & Osborne, 2014; Mujtaba & Reiss, 2014). This also approximated a continuous scale to a greater degree than any individual item considered alone, which helped meet the assumptions of (linear) predictive modelling (e.g. the outcome has a normal distribution, given the various predictors) (Bartholomew, Steele, Moustaki, & Galbraith, 2008; Cohen & Cohen, 1984).

While any individual item could form an outcome in itself in order to provide greater insight, and particular groups could be considered (e.g. those who responded with any degree of agreement, those who responded with strong agreement, etc.), these areas were outside the scope of the current research (and would entail an extensive amount of replication, further exploration of categorisation or grouping of students, etc.).

Self-concept beliefs

Historical research into self-concept beliefs initially focused on structural features (Marsh, 1990; Marsh, Byrne, & Shavelson, 1988), given a relatively broad initial conceptualisation (Shavelson, Hubner, & Stanton, 1976), and it perhaps remained unclear and/or unquestioned why self-concept was measured in particular ways or what aspects were necessarily integral. For example, initially, self-concept beliefs and interest/enjoyment were combined (Arens, Seeshing Yeung, Craven, & Hasselhorn, 2011; Marsh, Craven, & Debus, 1999). Contemporary research still potentially involves disconnections between theory, conceptualisations, and operational measurement. For example, research has proposed that peer-comparisons are influences on self-concept beliefs, but has sometimes included implicit expressions of peer-comparisons (e.g. 'Math is harder for me than for many of my classmates') as inherent aspects of a measure of self-concept (Marsh, Abduljabbar, et al., 2015). Essentially, in some cases, there may be less distinction between potential antecedents and potential expressions of self-concept beliefs.

Prior instruments and measures of self-concept were then reviewed (e.g. Martin & Mullis, 2013; OECD, 2009a, 2013). The various items were categorised and commonalities were identified (but potential antecedents were identified and considered separately): prior items broadly covered self-evaluation, general mastery experiences, mastery experiences of difficult work, and learning quickly (**Table 3**). These commonalities reflected aspects of existing theory (Bong & Clark, 1999; Bong & Skaalvik, 2003); higher abilities entailing less learning time have also been proposed by theories of learning (Carroll, 1989).

Accordingly, providing broad comparability with prior research, students' subject-level self-concept beliefs were measured through five items: 'I usually do well in science'; 'I have always been good at science'; 'I get good grades in science'; 'I understand even the most difficult science work'; and 'I learn things quickly in science'.

Self-efficacy beliefs

Someone's self-efficacy or confidence in their future capacities inherently requires contextually-dependent expressions (Bandura, 1997). For example, self-efficacy could be expressed as confidence to correctly answer particular tasks, confidence in gaining particular examination grades, or confidence in passing a particular course; self-efficacy as measured in PISA 2006 considered students' confidence in being able to undertake various non-curricular or everyday science-related tasks/activities (OECD, 2009a). General advice has been to measure self-efficacy on the same level as the outcome or area being investigated (Bandura, 1997; Bong, 1997; Pajares &

Miller, 1995). In the context of considering influences onto students' subject-level studying intentions, this entailed a subject-level expression of self-efficacy.

Accordingly, subject-level self-efficacy was measured through students' confidence expressed as their expected future attainment (i.e. 'What grade do you think you will be able to get at GCSE (or equivalent) science' and 'What grade do you think you would be able to get if you studied your best science subject at A-Level'). These indicators were also contextually-relevant as students likely require particular grades in order to enrol on A-Level or university courses. Prior research has similarly measured self-efficacy as expressions of future capabilities to gain course-specific attainment (Bong, 2001b; Lee, Lee, & Bong, 2014), although it remains possible that any number of other items could be developed.

Science grades

Students also recorded their current science grade and related information such as their science grade in the previous year and their average grade across all subjects. From these items, preliminary analysis highlighted that the students' current grade had the strongest association with science intentions (and with self-concept beliefs), and so was subsequently used within the final analysis. Students' previous grades (and other such information) were then used as background information during preliminary analysis, for example to predict the students' current grade in order to provide wider insight.

Two of the twelve participating schools requested that the grade information was measured through National Curriculum levels: the questionnaires were amended so that grades/levels were reported as free-text, which was then coded and categorised to be equivalent with the other reported grade data (detailed in **Appendix 4**). From the 685 students surveyed at Year 9, 186 students reported National Curriculum levels rather than alphabetical grades. Preliminary analysis suggested that the grades and levels associated with other reports but with some variability. For example, for Year 9 students (without inferring any missing values), the students' self-concept beliefs correlated slightly more strongly with reported grades (R = .445, p < .001) compared to National Curriculum levels (R = .372, p < .001). Conversely, self-efficacy beliefs (inherently expressed as expected alphabetical grades) correlated slightly more strongly with expressed current levels (R = .689, p < .001) compared to expressed current grades (R = .563, p < .001). However, any differences may have followed from the varying numbers of students involved and/or from the different students having different characteristics (and/or being at different schools).

Fundamentally, preliminary analysis produced similar predictive coefficients and significance values when considering only those students who reported alphabetical grades and when considering all students through the aggregated grade/level indicator, suggesting that the inclusion of those who reported National Curriculum levels was not notably influencing the various associations when considered within a wider context.

Regardless of these particular operational aspects, alphabetical grades may be inherently variable in implementation or interpretation across schools (outside of national examinations such as GCSE and A-Level grades), which unavoidably introduces variability or uncertainty. Additionally, self-reported grades can be under-reported, accurate, or over-reported (Robins & Beer, 2001; Gramzow, Elliot, Asher, & McGregor, 2003), but have generally been observed to have high correlations with actual grades (Kuncel, Credé, & Thomas, 2005).

During data collection, it was operationally unfeasible to collect information other than self-reports. For example, asking schools to provide ‘objective’ attainment records for their students would have been intrusive, eliminate anonymity (i.e. the questionnaires would have needed to ask for students’ names so that attainment records could be matched to their responses), and generally require further time and effort from teachers. In any event, for those students in Years 9, 10, and 11, any national attainment results would either be historical (e.g. Key Stage 2 tests taken in Year 6) or may not have been undertaken (e.g. GCSE or equivalent examinations), and so would not necessarily help to consider whether the students were tending towards under-confidence or over-confidence. Appearances of biases could follow from any unobserved changes in attainment over time; for example, someone might be generally accurate when reporting high self-concept beliefs but with low prior attainment (e.g. at Key Stage 2) if their attainment had subsequently increased in the meantime.

Preliminary analysis considered the accuracy/bias of students' self-concept beliefs compared to their current grades, but ultimately focused on the accuracy/bias of their task-level confidence and scores, given that these were more explicit (but contextualised) measures of self-reflection (**Section 5.3.4**).

Theorised influences on confidence

Various influences on, antecedents of, and/or sources of self-efficacy and self-concept beliefs have been identified or theorised (Bandura, 1997; Bong & Clark, 1999; Bong & Skaalvik, 2003). Measurement items were again informed by commonalities across various prior studies or instruments (e.g. Martin & Mullis, 2013; OECD, 2013; Usher & Pajares, 2009).

Accordingly, various potential influences on students' self-efficacy/self-concept beliefs were measured: mastery experiences (students' current science grade, as above); vicarious experiences ('When I see how another student solves a science problem, I can see myself solving the problem in the same way'); social persuasions or praise (e.g. 'My science teacher tells me I am good at science'); (the absence of) anxiety (e.g. 'Science makes me confused and nervous', reverse-scored); subject-comparisons ('Science is harder for me than any other subject', reverse-scored); and peer-comparisons ('Science is harder for me than for many of my classmates', reverse-scored).

Some items were unavoidably measured through single-items given the constraints of the questionnaire length and given few items or precedents being used in prior research. For example, for brevity and comparability with prior research, single-item indicators representing subject-comparisons and peer-comparisons were applied and negatively-phrased (Martin & Mullis, 2013). As before, all items were then coded so that higher scores reflected positive aspects/beliefs (e.g. science being easier than other subjects). Reassuringly, single-items have indeed been established as acceptable indicators when compared to multiple-item factors (Gogol, et al., 2014). Applying single-items increases the reliance on the particular phrasing, however; for example, the indicator of subject-comparisons and peer-comparisons considered relative difficulty rather than relative attainment comparisons. An extensive amount of research has focused on inferring the (implicit) influence of peer-comparisons through modelling students' own attainment and group-average levels of attainment (Marsh, Abduljabbar, et al., 2015; Nagengast & Marsh, 2011). Other research has, however, highlighted the benefit of considering (explicit) indicators of students' peer-comparisons through questionnaire items (Huguet, et al., 2009; Thijs, Verkuyten, & Helmond, 2010).

Anxiety and praise were measured with more extensive sets of items, given their prevalence in prior research (Martin & Mullis, 2013; OECD, 2013). However, conceptually, 'anxiety' may form one aspect of the 'costs' associated with studying a subject (Flake, Barron, Hulleman, McCoach, & Welsh, 2015). For the considered students, factor analysis confirmed that the anxiety and cost items indeed formed two factors, although acceptable reliability could still be observed across the aggregated anxiety and cost items (e.g. α = .826 across Years 9-11). Nevertheless, anxiety and costs were kept separate for direct comparability with prior research.

Theorised influences on intentions

The theorised influences on students' intentions and choices from the expectancy-value model have been frequently explored across various prior studies (Bøe & Henriksen, 2015). As before, measurement items were informed by commonalities across various prior studies or instruments (e.g. Conley, 2012; Martin & Mullis, 2013; OECD, 2013; Trautwein, et al., 2012; Wigfield & Eccles, 2000).

Students' science interest value reflected their inherent or intrinsic interest (e.g. 'I am interested in the things I learn in science') and enjoyment (e.g. 'I enjoy learning science') in studying science, and in science considered in general (i.e. 'I like science', interpretable as science at school, science as a wider field, etc.).

Utility value aimed to reflect the indirect or extrinsic benefits, importance, or value associated with science or studying science. This included the potential benefits of science for other areas of study (e.g. 'I need science to learn other school subjects', 'Science is an important subject for me because I need it for what I want to study later on') and for future employment (e.g. 'I will learn many things in science that will help me get a job', 'Learning science is worthwhile for me because it will improve my career prospects'). Exploratory factor analysis (across all items) highlighted potential associations between some utility value items (e.g. 'I need to do well in science to get the job I want') and items intending to directly measure science intentions. Nevertheless, further exploratory and confirmatory factor analysis highlighted that the intentions and utility items (when considered together) indeed formed two factors.

Students' personal value of science reflected the importance of science to their own identity. Personal value was considered as personal importance, as an inherent aspect of personal identity, and as a means to convey personal identity to other people (i.e. 'Science is important to me personally', 'Thinking scientifically is an important part of who I am', 'Being able to do science helps me show other people who I am'), and accordingly directly linked with theoretical conceptualisations of 'identity' within science (Carlone & Johnson, 2007). Potentially-similar items for measuring utility value (e.g. 'It is important to do well in science') (Martin & Mullis, 2013) did indeed load onto the utility value factor and not the personal value factor.

The cost value associated with science covered time (e.g. 'I have to invest a lot of time to get good grades in science'), lost opportunities (e.g. 'Success in science means that I need to give up other activities I enjoy'), and in general terms (e.g. 'I have to give up a lot to do well in science'). Conceptually, costs can broadly encompass effort and time, demands and restrictions from other areas of life, sacrifice and the loss of other alternatives, and negative emotions such as anxiety and stress (Flake, Barron, Hulleman, McCoach, & Welsh, 2015; Wigfield & Eccles, 2000). The items measuring anxiety and cost nevertheless formed separate factors for the considered students.

Some students completed a questionnaire with only two items for each of the personal value and cost value factors, however, since the last iteration of the questionnaire was (perhaps overly) reduced in length to help teachers administer it. In retrospect, other factors could have been more-easily reduced and the importance of the personal value factor was less immediately clear, given prior research focusing on utility value and interest value. Nevertheless, the full items were available for the majority of students, and prior studies have indeed measured factors such as cost with only two items (Conley, 2012; Trautwein, et al., 2012).

Other potential influences on intentions

While the expectancy-value model recognises that someone's context and other people may be influential on intentions/choices, further factors have not been consistently modelled, given that wider influences are theorised to be mainly mediated by someone's confidence and their various ('subjective task value') attitudes (Wigfield & Eccles, 2000). Prior research has nevertheless highlighted the direct influence of other people onto students' choices (Mujtaba & Reiss, 2014; Sjaastad, 2012). Accordingly, further potential influences on students' intentions were measured, including the students' perceptions of their teacher and/or learning context (Martin & Mullis, 2013), any potential influences of friends and parents (OECD, 2013), and students' notions of perceived control or effort (OECD, 2013). Items were broadly phrased as per earlier research for comparability.

Students' perceptions of their teacher and/or learning context covered their affective perceptions (e.g. 'My Science teacher is easy to understand', 'My science teacher gives me interesting things to do') together with any experiences of formative feedback/guidance (e.g. 'My Science teacher gives me feedback on my strengths and weaknesses in science', 'My Science teacher tells me what I need to do to become better in science') (Martin & Mullis, 2013). Factor analysis confirmed that these various items formed only one factor.

A further factor was formed to cover the provision of science careers advice, events, and guidance from teachers or otherwise provided by the school (e.g. 'My science teacher tells me about careers and jobs in science').

Implicit influences or 'subjective-norms' were considered in relation to the students' friends (i.e. 'Most of my friends do well in science', 'Most of my friends work hard at science', 'My friends enjoy taking science tests') and parents ('My parents believe it's important for me to study science', 'My parents believe that science is important for my career', 'My parents like science'). Factor analysis confirmed separate factors for friends and parents although acceptable reliability was also observed across the combined items.

Someone's 'perceived control' was originally (historically) conceived as theoretically akin to self-efficacy but also considering the perceived ease or difficulty of the area (Ajzen, 1991, 2002). Perceived control was intended to be distinct from someone's beliefs of their 'locus of control', considered as whether outcomes followed from someone's own efforts/characteristics (internal causes) or wider (external) causes (Rotter, 1966). Similarly, perceived control did not necessarily consider whether someone believed that their abilities or other personal characteristics were fixed or changeable, and so were broadly under their own control or not in terms of potential development (Dweck, 2000; Dweck & Leggett, 1988). Contemporary measurement, however, has broadly considered someone's perceived control as whether their personal efforts can lead to success, which perhaps implicitly considers an internal locus of control (OECD, 2013; Pintrich, Smith, Garcia, & Mckeachie, 1993; Pintrich, Smith, Garcia, & Wilbert, 1991).

Perceived control was nevertheless considered as a potential motivational factor, even though the underlying concept appeared less clear. For example, it was possible to hypothesise that believing that science abilities cannot be changed, outcomes are outside of personal control, and/or effort is futile, might entail lower intentions to study science further, and/or somehow link with under-confidence. Reviewed across various studies, believing that personal abilities could be changed has associated with higher attainment (to a small extent) and with various other beneficial motivational beliefs, such as motivations to learn and master work, and expectations that outcomes could be achieved (Burnette, O'Boyle, VanEpps, Pollack, & Finkel, 2013). For students at the start of secondary school in the United States, for example, believing that personal science abilities could be changed associated with higher science self-efficacy, and boys reported stronger beliefs compared to girls (Chen & Pajares, 2010). Perceived control for learning, considered akin to an internal locus of control, has also associated with high attainment and self-efficacy for university students (Credé & Phillips, 2011).

Factor analysis highlighted that items measuring the notion of ability being changeable (e.g. 'I can improve my ability in science') (Dweck, 2000) formed one factor together with those items measuring perceived control in general terms, covering whether personal effort could lead to success (e.g. 'If I put in enough effort I can succeed in science') and whether success would follow from individual efforts (e.g. 'Whether or not I do well in science is completely up to me') (OECD, 2013). However, new (negatively-phrased) items considering perceived control for attainment (e.g. 'I do badly in science whether or not I study for my exams') formed a separate factor. Indicators of reliability were also higher for these two separate factors than across the aggregated items.

As further potential motivational influences, students' orientations towards mastering learning (referred to as a 'mastery' orientation) and/or performing better than other students (referred to as a 'performance' orientation) were also measured (Ames, 1992; Dweck & Leggett, 1988; Elliot, 1999; Elliot & McGregor, 2001; Elliot & Thrash, 2001). Prior conceptualisations and measurement of these goal orientations has perhaps tended towards re-phrasing the same (singular) underlying idea across multiple items (e.g. 'I am striving to demonstrate my ability relative to others in this class', 'I am motivated by the thought of outperforming my peers, 'It is important to me to do well compared to others in this class') (Elliot, 1999). Given that the underlying ideas appeared to form relatively clear and distinct concepts, and for brevity, new single-items were applied in the questionnaire ('I aim to understand and learn the material in science' and 'I aim to perform better than other students in science').

Fundamentally, the inclusion of these further items/factors could potentially extend understanding and highlight their relevance to science intentions and/or confidence accuracy/bias, or could conversely highlight their irrelevance.

Self-regulation and studying strategies

Self-regulation for studying can be measured in various ways. Various questionnaire items have been developed, although arising more from prior empirical studies than deriving from theoretical cyclical models of self-regulation (Pintrich, 2004; Pintrich, Smith, Garcia, & Mckeachie, 1993; Pintrich, Smith, Garcia, & Wilbert, 1991; Zimmerman & Martinez-Pons, 1986, 1988). Various items/factors within different instruments have been found to be broadly equivalent (Muis, Winne, & Jamieson-Noel, 2007). Other approaches to measure self-regulation have facilitated students to talk through their experiences and processes of undertaking tasks (Armstrong, Wallace, & Chang, 2008) or have considered recordings or observations (coded by researchers) of how students work in practice (Lippmann Kung & Linder, 2007). When multiple methodological approaches have been considered for the same students, results have variously been similar across the approaches (Schellings, 2011; Schellings, van Hout-Wolters, Veenman, & Meijer, 2013) or different across the approaches (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007; Jacobse & Harskamp, 2012). Within the constraints of the research underlying this thesis, however, it was unfeasible to apply multiple approaches in addition to the questionnaire.

The self-regulation for studying factor ('meta-cognitive self-regulation') from the Motivated Strategies for Learning Questionnaire (MSLQ) was selected for the questionnaire, given that the overall instrument was broadly based on the social-cognitive model, had been applied within various prior research to aid contextualisation, and that the relevant factor was relatively brief (Pintrich, Smith, Garcia, & Mckeachie, 1993; Pintrich, Smith, Garcia, & Wilbert, 1991). The factor covered various areas including the setting of goals (e.g. 'When I study for my science class, I set goals for myself in order to direct my activities in each study period'), adapting and applying different studying approaches (e.g. 'I try to change the way I study in order to fit the science course requirements and the teacher's teaching style'), and reflection and/or monitoring (e.g. 'I ask myself questions to make sure I understand the material I have been studying earlier in science class'). This broadly covered elements within theorised cycles of self-regulation (Zimmerman, 2000), although without explicitly asking whether a cycle or process occurs.

The MSLQ items were used directly for the measure of self-regulated studying, in order to be more comparable with prior studies. However, some items may not necessarily have been ideal though using conditional phrasing (e.g. 'If science course materials are difficult to understand, I change the way I approach the material'), which may introduce uncertainty (e.g. some students may not find science materials difficult to understand) and/or entail that agreement or disagreement may have different meanings for different students. Students may have attempted to interpret what the item was asking, given that the exact phrasing may not have completely applied to them.

Further measures were also included for the study strategies of 'controlling' or organising learning (e.g. 'When I study science, I start by working out exactly what I need to learn', 'When I study science, I try to figure out which concepts I still have not understood properly'), 'memorisation' or rehearsal (e.g. 'When I study for a science test, I learn as much as I can off by heart', 'In order to remember the method for solving a science problem, I go through examples again and again'), and 'elaborating' materials and ideas (e.g. 'When I study for a science test, I try to understand new concepts by relating them to things I already know', 'When I study science, I try to relate the work to things I have learnt in other subjects'). Such factors have been similarly measured in the MSLQ (Pintrich, Smith, Garcia, & Wilbert, 1991) and international studies (OECD, 2013), and were phrased for comparability with both.

These studying strategies provided further potential indicators of benefits or detriments that might associate with students' confidence accuracy/bias. Across various studies with university students, strategies of control, memorisation, and elaboration all associated with higher attainment, with elaboration having the highest association (Credé & Phillips, 2011; Richardson, Abraham, & Bond, 2012). For secondary school students, however, memorisation strategies have associated with lower task-scores in mathematics, and control strategies have associated with higher task-scores, while elaboration has had no association, in PISA 2000 (Chiu, Chow, & Mcbride-Chang, 2007) and in PISA 2012 (Echazarra, Salinas, Méndez, Denis, & Rech, 2016). Similarly, for secondary school students in Germany (outside of PISA), memorisation associated with lower attainment in mathematics, controlling for the students' interest in mathematics and other factors, while elaboration had no association (Köller, 2001).

Study strategies are nevertheless distinct from the concepts of 'surface' and 'deep' learning (Biggs, 1993; Entwistle & McCune, 2004); memorisation does not necessarily imply or equate to surface learning (i.e. only learning the minimum that is sufficient to pass), nor does elaboration equate to deep learning (i.e. maximising learning). The ideas of surface and deep learning consider motivations or orientations towards learning rather than the particular strategies undertaken when learning (Pintrich, Smith, Garcia, & Wilbert, 1991).

Some of the self-regulated studying items (e.g. 'I try to think through a science topic and decide what I am supposed to learn from it rather than just reading it over when studying for science') might conceptually overlap with control strategies (e.g. 'When I study for a science test, I try to work out what the most important parts to learn are'). Exploratory factor analysis highlighted that the self-regulated learning strategies and the other learning strategies could overlap to varying degrees (i.e. common factors could emerge across the various strategy items) but the results varied depending on which students were considered (e.g. individual year groups and/or across Years 9-11) and whether all questionnaire items were considered or whether only the studying strategy items were considered. Given that confirmatory factor analysis highlighted single-factors with acceptable reliability for each factor considered alone, the theorised items/factors were used for direct comparability with existing research.

Students' background

Students' self-reported background was also recorded. Specifically: their gender; their background/ethnicity; the highest level of education completed by the students' mother and father (or equivalent guardians); the number of books at home; and whether either parent/guardian worked in any job or area related to science (as interpreted by the student).

These indicators balanced including those measured within comparable prior studies against brevity and areas that students may be able to answer (e.g. Archer, Dawson, DeWitt, Seakins, & Wong, 2015; OECD, 2009a). It was less feasible to include complex measures of parental occupation and/or to solicit free-text descriptions that would then require coding and classification. For example, family background is often considered through indicators of 'socio-economic status', which is often considered through classifying occupations. National surveys often apply the 'National Statistics Socio-economic Classification' (NS-SEC) scheme, which considers someone's occupation and employment status, managerial responsibilities, and workplace size (Rose & O'Reilly, 1998; Rose, Pevalin, & O'Reilly, 2005). However, such questions are unfeasible for students to answer. Additionally, the 'economic' aspect of this 'socio-economic classification' appears to remain implicit, so it perhaps remains unclear what is or should be measured; students might again be less likely to know the precise details of their parents' income, and/or find the question intrusive. The piloting indeed highlighted (anecdotally from free-text responses) that some students found questions about their parents intrusive and they were not clear why they were asked.

Preliminary analysis of PISA 2006 highlighted that indicators of parents' occupational levels were not significantly predictive of students' science intentions, when considering parental education, parents working in science or not, and the students' own attitudes and beliefs. Similarly, prior research has associated students' aspirations more with parental education rather than parent occupation (Davies, Qiu, & Davies, 2014). To facilitate the data collection in the 2014/2015 survey, students were then only asked about their parents' levels of education and whether either parent worked within science. Not considering hierarchies of parental occupation may be considered problematic for some research fields, but appeared to be less contextually relevant here (i.e. parents working within science or not appeared to have more potential relevance); nevertheless, it may be beneficial to explore all such areas within future studies.

Further areas such as students' science set were recorded by the questionnaire (i.e. top, middle, or bottom, if setting was used) but these had no predictive association with science intentions once students' wider attitudes and beliefs were also modelled (specifically, setting was completely mediated by students' self-concept and self-efficacy beliefs). For brevity, such areas were then not included within the final analysis; since the analysis focused on accuracy/bias groups/clusters, further sub-division by science sets could not be considered (i.e. these would be too small for predictive modelling).

Section 5.3.4: Measuring students' confidence biases

Evaluating someone's confidence accuracy/bias essentially requires conceptually-equivalent measures of confidence and attainment, so that the two can be plausibly compared (**Section 3.4**). The most explicit comparisons have involved attainment tasks paired with expressions of confidence, so that self-reflective confidence accuracy/bias can be directly considered (Chen, 2003; Pajares & Graham, 1999). In contrast, comparisons of attainment and confidence through PISA, for example, can only form implicit or potentially artificial indicators of accuracy/bias (Cheema & Skultety, 2016; Chiu & Klassen, 2010).

Potential attainment tasks for the questionnaire were considered from a variety of sources, including PISA, TIMSS, discontinued national Key Stage 3 (KS3) tests, and previous examination papers for GCSE and A-Level tests from various providers (e.g. AQA, Edexcel, OCR, etc.). All had been nationally or internationally validated as reliable indicators of performance through various processes. Tasks from TIMSS were then selected due their strong contextual relevance (i.e. being designed to cover curricula areas) and their efficiency in measurement (Foy, Arora, & Stanco, 2013). Additionally, TIMSS tasks have been successfully used in prior research to consider confidence accuracy/bias (Chen, 2003; Chen & Zimmerman, 2007; Seidel, 2006). Alternately, tasks from PISA were relatively lengthy and often involved multi-stage tasks, and were less contextually-relevant through considering applied rather than curricular areas (OECD, 2006b). Tasks from past examination papers or legacy KS3 tests were also relatively lengthy, and although they potentially offered strong contextual relevance they were potentially less accessible to different ages of students.

Tasks from TIMSS have been internationally validated through extensive processes, and were designed to broadly cover curricula areas from the participating countries, including England (Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009). As a precaution, the content of the tasks were verified against the (Key Stage 3) National Curriculum (Department for Education, 2013, 2014). For example, the 'what is a compound' task (S042306) operated within the 'atoms, elements and compounds' National Curriculum area, where students are required to understand the differences between atoms, elements, and compounds; the 'parachute jumper' task (S032141) operated within the 'motion and forces' National Curriculum area, where students need to consider and understand balanced and unbalanced forces.

Reliability (i.e. consistency) and validity (i.e. scope of content coverage) were both likely to increase with the number of tasks (Cohen, Manion, & Morrison, 2007; Cronbach, 1951). It was unfeasible to assign students a comprehensive examination, however. The overall purpose was not to 'definitively' measure attainment, but to measure accuracy/bias through pairs of tasks and confidence-ratings (although reliability/validity would nevertheless be improved through considering more task/confidence pairs). Prior research into accuracy/bias has considered variable numbers of paired items, for example fifteen (Chen, 2003; Chen & Zimmerman, 2007). Including ten tasks was feasible, balanced against the other areas of the questionnaire and its overall length. Calculations of accuracy/bias then covered similar numbers of items to longer attitudinal scales, and covered more items than many factors (such as self-concept beliefs).

The tasks were selected to cover a range of curricular areas, including photosynthesis, atomic structures, changes of state, electricity and current, and various other areas; these covered biology, chemistry, and physics. The selected tasks mainly used a four-item multiple-choice format but also involved students writing free-text responses, which were subsequently coded using the TIMSS schemes (Foy, Arora, & Stanco, 2013).

The tasks were selected to be potentially accessible to younger and older students. Given published task-level results and analysis of TIMSS 2011 data, tasks were selected with a range of likely 'difficulties' (i.e. inferred from the average proportion correct in TIMSS 2011 across students in England), but with a balanced overall average (i.e. averaging close to 50% answered correctly across the selected tasks) (Foy, Arora, & Stanco, 2013). Essentially, the selected tasks were intended to be (considered together) neither too easy nor too difficult, and so to be relatively valid for different ages. Subsequently, across the surveyed students in 2014/2015, the proportions of correct answers per item appeared broadly similar to those observed in TIMSS 2011 (Foy, Arora, & Stanco, 2013). Even for Year 11 students, on average, there did not appear to be 'ceiling' effects with all students essentially answering items correctly. These particular item-level details are shown in **Appendix 5**.

In preliminary analysis, as in PISA and TIMSS, item-response models were considered in order to estimate students' overall task-scores: such models estimated the varying 'difficulty' of each item while concurrently estimating the students' performance (de Ayala, 2009; Rabe-Hesketh, Skrondal, & Pickles, 2004; Zheng & Rabe-Hesketh, 2007). Preliminary analysis highlighted that, for example, one-parameter-logistic item-response model estimates of performance strongly associated with average proportion-correct task-scores (e.g. $R = .903$, $p < .001$, across Years 9-11 without inferring any missing values). The various item-response model parameters and estimates depended on the considered students, however, so that the statistical sophistication of estimating item difficulty and performance per academic year was potentially undermined by any uncertainty arising from the lower numbers of modelled students for each year. Ultimately, sophisticated measures were not required to compare someone being correct or not against being confident or not in their answer. Therefore, students' task-scores were measured as the average proportion of correct answers.

Each task was paired with a confidence rating (i.e. 'How confident are you that you solved this correctly?'), providing a retrospective self-reflective evaluation of the students' answer. Various other self-evaluative or 'meta-cognitive' questions can be asked immediately after a task, for example covering the familiarity, interest, perceived difficulty, and applied effort and time related to solving the task (Efklides, Kourkoulou, Mitsiou, & Ziliaskopoulou, 2006). Only retrospective task-confidence was considered for brevity.

Measures of confidence given before and after undertaking tasks conceptually differ (i.e. respectively reflecting extremely contextualised self-efficacy and self-concept beliefs to some extent) and may be formed in different ways. For example, someone may consider features of the task in order to express their prospective confidence in being able to successfully solve it, and/or (perhaps if they were unfamiliar with the task) they may also generalise from their wider subject-level confidence beliefs. Alternately, confidence retrospectively expressed after undertaking tasks allows someone to self-reflect on their problem-solving processes and their answers. Accordingly, and as highlighted in prior studies, confidence expressed after undertaking tasks may be inherently more accurate than expressed before undertaking tasks, although under-confidence and over-confidence biases have nevertheless still been observed (Ackerman & Wolman, 2007; Boekaerts & Rozendaal, 2010; Erickson & Heit, 2015).

In the last version of the questionnaire, two tasks were removed in order to substantially reduce the length to help teachers administer the questionnaire and for students to complete the questionnaire within limited time periods. The various task-scores and accuracy/bias indicators were accordingly calculated depending on the questionnaire version (i.e. out of eight or ten items/pairs).

Indicators of task-level confidence biases

Various indicators can be calculated to quantify overall accuracy and/or bias in confidence (Lichtenstein, Fischhoff, & Phillips, 1982; Yaniv, Yates, & Smith, 1991; Yates, 1990).

Following earlier studies (Chen, 2003; Pajares & Graham, 1999), a simple difference-score was calculated between the students' average task-confidence and task-score, equalised to the same scales. This measured the ‘absolute’ degree of accuracy/bias, from under-confidence through accuracy through to over-confidence. Preliminary analysis also considered regression-residual indicators, calculated separately for each academic year, for task-level confidence accuracy/bias (i.e. ‘relative’ under-confidence or over-confidence via comparing someone’s reported task-confidence against the predicted task-confidence that would be expected given their task-score and given the task-score/task-confidence association across the students for that academic year) and for subject-level self-concept accuracy/bias (i.e. ‘relative’ under-confidence or over-confidence in reported science self-concept, compared to the predicted self-concept given the students’ current science grade and given the other students for that academic year).

The task-level difference-score and regression-residual indicators closely associated (e.g. $R = .777$, $p < .001$, without accounting for missing values, across Years 9-11). However, the regression-residual indicator of self-concept accuracy/bias had less association with the task-level difference-score ($R = .198$, $p < .001$) and the task-level regression-residual indicators ($R = .358$, $p < .001$, both without accounting for missing values, across Years 9-11). The other associations between the task-level and subject-level indicators might suggest a higher correspondence between accuracy/bias across both levels. Without accounting for missing values and considered across Years 9-11, moderate (and relatively similar) associations were observed between students’ task-scores and current grades ($R = .515$, $p < .001$), task-confidence and self-concept beliefs ($R = .557$, $p < .001$), task-scores and task-confidence ($R = .521$, $p < .001$), and current grades and self-concept beliefs ($R = .489$, $p < .001$).

Preliminary analysis highlighted that for the 2014/2015 survey, self-concept had little predictive association with the students’ science intentions, when controlling for the students’ other attitudes and beliefs (in contrast to the PISA 2006 preliminary results). Considering self-concept accuracy/bias may have then been less meaningful in context, for this new sample of students. Alternately, considering self-efficacy accuracy/bias (i.e. confidence, expressed as expected grades, compared to actual grades) would not have been possible without a longitudinal design. Fundamentally, in contrast to comparing self-concept beliefs and current grades, the task-level comparisons considered self-reflective accuracy/bias with potentially higher validity (i.e. task-scores and task-confidence clearly considered the same area) and reliability (i.e. multiple pairs of task-scores/confidence-ratings were considered). On the task-level, it was also less meaningful to consider 'relative' (regression-residual) accuracy/bias when 'absolute' (difference-score) accuracy/bias could be considered. Accordingly, the difference-score indicator of task-level accuracy/bias was focused on within the final analysis.

Various other indicators of accuracy can be considered through comparing binary measures of confidence against correctness. For these approaches, the scalar/categorical confidence-ratings were coded as either being confident (i.e. 'Very confident' or 'Confident') or not confident (i.e. 'Not very confident' or 'Not at all confident'). Indicators of 'sensitivity' and 'specificity' were calculated in order to provide additional insight; sensitivity and specificity have a long history of applications within statistics and medical/clinical diagnosis, and both provide indicators of accuracy (Schraw, Kuch, & Gutierrez, 2013; Schraw, Kuch, Gutierrez, & Richmond, 2014; Yule, 1912).

Sensitivity provides an indicator of when someone knows that they have answered correctly (i.e. being 'confident' when they have the right answer) while specificity provides an indicator of when someone knows that they have the wrong answer (i.e. being 'not confident' when they have the wrong answer) (Schraw, Kuch, & Gutierrez, 2013; Yule, 1912). A complementary indicator of 'simple-matching' was also calculated, representing the combined proportions of both knowing when answers are right and knowing when answers are wrong (Schraw, Kuch, & Gutierrez, 2013; Yule, 1912). The calculation methods/formulae are reproduced in **Appendix 6**. Given that the sensitivity, specificity, and simple-matching indicators provided scales of accuracy (and not biases towards under-confidence and over-confidence), they were not considered as means for grouping students. This remained a potential area for future exploration.

Section 6: Analytical approaches

The research presented in this thesis considered three main areas of enquiry (**Section 4**) through two sets of survey data (**Section 5**). Given the quantitative design, the research questions were addressed in the following ways.

The first research question, considering which attitudes and motivational beliefs were the most relevant influences on students' science intentions and choices, entailed undertaking predictive modelling of the students' reported science intentions and determining which predictors had the highest relative magnitudes. The relevant results and discussion are covered in **Section 7**.

The second research question, considering whether students with different degrees of confidence accuracy/bias exhibited different science intentions, attitudes, and beliefs, entailed categorising students according to their different degrees of confidence accuracy/bias (via groups and clusters), and comparing their expressed beliefs across the various groups and clusters, especially considering students' intentions and any key predictors identified from the first research question. The relevant results and discussion are covered in **Section 8**. Differences across groups/clusters were determined via analysis of variance tests with Bonferroni post-hoc tests (reproduced via sample-weighted general linear models for PISA 2006).

The third research question, considering whether students with different degrees of confidence accuracy/bias considered their science intentions in different ways, entailed applying predictive models of students' intentions per group/cluster and considering whether the patterns of coefficients differed across different groups/clusters. The relevant results and discussion are covered in **Section 9**. Significant differences in coefficient magnitudes across groups (or clusters) were identified through interaction models: two groups were modelled together (e.g. under-confident students and accurate students); the various items/factors were included as predictors, together with a group membership indicator (e.g. accurate = 0 or 1) and the interactions between the group membership indicator and each of the other predictors. The significance associated with the interaction terms then highlighted differences in coefficient magnitude across the two groups being considered.

Within each survey, the three research questions considered common sets of items/factors: essentially, there were no items/factors considered in the second and third research questions that were not considered in the first research question.

Before the various results are described and discussed, the following sections describe the common methodological and analytical approaches underlying these three research areas, including highlighting potential methodological limitations which also informed the selection of specific approaches.

Students do not necessarily respond to every questionnaire item, and expectation-maximisation was applied across both surveys to estimate missing values/responses and hence to maximise the number of considered students, otherwise undertaking predictive modelling could be unfeasible for smaller groups or clusters of students (**Section 6.1**).

Predictive models can also be undertaken in various ways, and multi-level models were more suited to the structure of students within schools than single-level models (**Section 6.2**).

Students can also be grouped or clustered in various ways in order to consider their confidence accuracy/bias, such as via researchers defining specific groups (Gonida & Leondari, 2011), algorithmic hierarchical clustering (Sáinz & Upadyaya, 2016), algorithmic optimisation clustering (Rytkönen, Aunola, & Nurmi, 2007), and model-based clustering (Seidel, 2006). The formation of groups, covering the conceptual ideas of under-confidence, accuracy, and over-confidence, is described in **Section 6.3**. Additionally, the model-based clustering approach of latent-profile analysis offered various advantages such as helping to identify optimal numbers of clusters while avoiding potential difficulties inherent to algorithmic and optimisation clustering, and was used to identify various clusters of students (**Section 6.3**).

Additionally, given the power/sample size calculations underlying the research design (**Section 5.1**), it was necessary to aggregate the 2014/2015 survey sample across academic years (i.e. Years 9, 10, and 11). **Section 6.4** describes how various approaches were applied to still ensure meaningful results, such as controlling for the students' academic year within predictive models.

As a wider point, across the various analyses, $p < .05$ was used as the criterion to denote statistical significance (Cohen, 1992). Considering larger numbers of students, such as in PISA studies, may allow smaller differences or coefficients to be revealed with statistical significance; accordingly, magnitudes of differences and coefficients were considered in order to help determine whether results were meaningful (Cohen, 1988, 1992).

Section 6.1: Missing responses

Students may have numerous reasons for not answering particular questions. They may not know the answer or may not understand the question, for example, or they may not have sufficient time or interest to respond. Within educational research, any set of collected data is likely to unavoidably contain missing responses for various items.

Forming factors by aggregating multiple items can help mitigate the impact of any missing responses (Bartholomew, Steele, Moustaki, & Galbraith, 2008; Bollen, 2002). A factor-score can be calculated as the average of five items, for example, even if some of these items may have been left blank by some students, ensuring that a factor-score is available for most (if not all) students. Missing responses may be more problematic for single-item indicators. Educational research often uses single-items as indicators of personal characteristics or background, for example of gender or of ethnicity. Additionally, single-items may be used as indicators of wider conceptual or theoretical factors. For example, parental occupations may be considered to reflect some form of status or categorisation (e.g. 'social class' or 'socio-economic classification') (Rose & O'Reilly, 1998), while parental education may be considered to reflect implicit dispositions within families (e.g. 'habitus') that may influence students' educational choices (Bourdieu, 1984). Similarly, indicators of the number of books at home, and/or other material assets, may also be considered to reflect family income or wealth (OECD, 2009a) and/or some form of cultural engagement or knowledge ('cultural capital') (Bourdieu, 1984). Problematically, students may not always answer such items.

For PISA 2006, relatively few missing values were present for the students' attitudes and motivational beliefs measured through the OECD factor-scores (e.g. a maximum of 4.1% missing for students' self-concept beliefs). Across the other considered items/factors, however, the most missing responses were observed for the students' reports of the educational level of their mother (9.9%) and father (15.2%).

For the 2014/2015 survey, numerous missing responses were present. Considered across Years 9-11, most missing responses occurred for the students' reports of the educational level of their mother (30.9% missing) and father (34.1% missing). Given that these items occurred on the first page of the questionnaire, and that the surrounding items had lower proportions of missing responses, it is perhaps plausible to assume that students simply did not know the particular details of their parents' education (or felt less comfortable in answering the question, or many other reasons may have been relevant; it remains unclear whether students could or would answer simpler questions such as whether their parents/guardians attended university or not). The various missing responses entailed that some missing values were present even for factors formed through aggregates of multiple items, including self-concept (2.2% missing across Years 9-11), self-efficacy (4.4%), interest value (11.2%), utility value (12.0%), personal value (14.0%), and cost value (14.2%). Higher proportions of missing responses were present for the students' reported studying strategies, which were measured at the end of the questionnaire, specifically the students' self-regulatory studying strategies (20.6% missing across Years 9-11) and control (21.9%), memorisation (21.3%), and elaboration (22.5%) strategies.

Predictive modelling often only considers students with responses/values for every modelled item/factor ('listwise deletion', the default within most statistical software). Only considering students with responses for every modelled item/factor (detailed in **Section 5**) would notably reduce the numbers of considered students within PISA 2006 (i.e. only considering 3860 of 4935 students when predicting science intentions) and the 2014/2015 survey (i.e. only considering 571 of 1523 students when predicting science intentions across Years 9, 10, and 11); these modelled students may then not necessarily reflect the wider sample and/or population. Additionally, lower numbers of students may reduce the statistical power of any tests (e.g. whether coefficients can be determined to be statistically-significantly different to zero or not). For the 2014/2015 survey, given the power/sample size calculations underlying the research design (**Section 5.1**), it was necessary to aggregate across academic years (i.e. Years 9, 10, and 11) and to maximise the number of students (i.e. not to only consider those with responses for every considered item/factor). For both surveys, maximising the number of students would help ensure the feasibility of modelling smaller groups or clusters, for example when undertaking predictive modelling for under-confident, accurate, and over-confident groups separately.

Additionally, only considering students with responses for every item/factor assumes that missing responses are 'missing completely at random' (Rubin, 1976). Missing responses across the various items/factors considered together did not appear to follow this assumption within PISA 2006 (Little's test: χ^2 (1677) = 2717.750, $p < .001$) and within the 2014/2015 survey (Little's test: χ^2 (4800) = 5792.674, $p < .001$) (IBM, 2014; Little, 1988). The missing responses could therefore either be 'missing at random' or 'missing but not at random' using formal terminology; it was impossible to determine either way without knowing the actual magnitudes of the missing values themselves (Rubin, 1976). Fundamentally, only considering the students who responded for every item/factor appeared not to be ideal.

Alternately, missing responses can be replaced with estimated values in order to allow most if not all students to be analysed. Applying full-information maximum-likelihood via expectation-maximisation or applying multiple-imputation to estimate missing responses are considered to be the best contemporary approaches to handling missing responses (Peugh & Enders, 2004; Rubin, 1996). Both approaches assume joint normality of all the modelled items/factors and that any missing responses are 'missing at random' (i.e. missing responses can be predicted by other observed items/factors, rather than assuming/requiring that missing responses are 'missing completely at random') (IBM, 2014; Peugh & Enders, 2004; StataCorp, 2013a). Specifically, full-information maximum-likelihood via expectation-maximisation estimates population parameters (i.e. means, covariance, correlations) that are likely to have produced the considered data; given these parameters, conditional expectations of any missing responses are then calculated, which are then used to further refine the parameters and re-estimate the missing responses in an iterative process (Dempster, Laird, & Rubin, 1977; IBM, 2014).

Regardless of the technical sophistication applied, someone's missing responses are unlikely to be perfectly estimated given the other available items/factors. Estimates of missing responses may also have less variability than might be present in 'real' data; standard errors in analysis may be slightly lower, and hence significance (low *p*-values) may be slightly more likely (Peugh & Enders, 2004). Multiple-imputation attempts to address this through applying expectation-maximisation but then saving multiple estimates ('plausible-values') of any missing responses, each with some randomly-added variation; analytical processes then combine results from across the multiple data sets to separate the sources of variation (Rubin, 1976, 1987). However, different plausible-values may be produced on different occasions, given the inherent randomness. Multiple-imputation also focuses on entire-sample models and assumes a consistent number of students across any analysis (Rubin, 1976, 1987); attempting to group students on plausible-values becomes complicated since the numbers of students per group could then vary across the various sets of data. Applying multiple-imputation then appeared to be less feasible, given that students needed to be grouped and clustered within the research design. Additionally, the available statistical software to undertake latent-profile analysis (Latent Gold) did not support multiple-imputation, although it allowed full-information maximum-likelihood estimation via expectation-maximisation (Vermunt & Magidson, 2013).

Within educational and motivational research, various studies have applied expectation-maximisation to directly estimate missing responses (Lee, Bong, & Kim, 2014) or applied expectation-maximisation through full-information maximum-likelihood via structural equation modelling (Trautwein, et al., 2012). Multiple-imputation appears to have been less frequently applied, although it remains an inherent aspect of PISA and TIMSS task-scores, which are estimates via multiple-imputation to infer across missing-by-design blocks of tasks (Foy, Arora, & Stanco, 2013; OECD, 2009a).

Fundamentally, applying expectation-maximisation appeared to be a feasible approach to handle missing responses, given that the analysis needed to form and analyse groups or clusters of students within wider samples while maximising the number of considered students. Potential influences on *p*-values were then considered/addressed through preliminary modelling with and without accounting for missing responses, and through considering magnitudes of 'effect' rather than significance alone. The preliminary analysis highlighted that models appeared sufficiently similar with and without estimating missing responses, when broadly similar numbers of students could be considered, suggesting that estimating missing responses did not appear to introduce issues.

Accordingly, for the final analysis, estimates of missing responses were produced through expectation-maximisation (IBM, 2014) using all available items/factors in each survey, plus additional items/factors that were available but were not included within the final analytical models (e.g. task-scores for mathematics and reading in PISA 2006). Essentially, models to estimate missing responses have been advised to be comprehensive, and to include available items/factors as additional input even if they are not included in the subsequent analytical models (Peugh & Enders, 2004). Estimates of missing responses could only be made for the continuous items/factors and not for binary/categorical items such as gender and whether either parent worked within science, although these items/factors were still included as input for the expectation-maximisation process (IBM, 2014). Missing factor-scores (plus the various single-item indicators when necessary) were estimated rather than the underlying items, which would have otherwise entailed unfeasibly extensive models.

As a precaution, only students' expressed beliefs (not including any estimates of missing responses) were used when considering confidence accuracy/bias (which unavoidably reduced the number of considered students to a small extent). If missing responses were estimated for self-concept beliefs or task-confidence using only task-scores, for example, then these estimated responses may be more likely to appear to be 'accurate' through regression-residual approaches. Nevertheless, preliminary analysis produced similar results when considering both situations, for example where latent-profile analysis resulted in similar accuracy/bias cluster profiles and proportions regardless of whether missing responses were estimated or not.

For consistency, the various final results (e.g. averages per cluster/group, correlations, predictive coefficients) include estimates of missing responses (unless highlighted otherwise), given that these were included within the data considered by the final predictive models.

Section 6.2: Predictive modelling

Predictive modelling can reveal the independent association between each predicting item/factor (e.g. interest, utility, etc.) and an outcome (e.g. students' science intentions). The various predictive associations (i.e. the estimated 'effect' of each item/factor on the students' science intentions, controlling for all the other predicting items/factors) can then be directly compared.

Predictive models via linear regression using ordinary-least-squares estimation involves numerous assumptions, specifically: independent observations (controlling for the predictors, the values of the outcome are independent across observations); linearity (the expected mean outcome is a linear function of the predictors); constant variance (the variance of the outcome, given the predictors, is constant, i.e. the variance of the outcome does not depend on the magnitude of the predictors); and normality (the outcome has a normal distribution, given the predictors) (Bartholomew, Steele, Moustaki, & Galbraith, 2008; Cohen & Cohen, 1984). Accordingly, the residual error (the differences between the observed outcome and the predicted outcome, given the model) is assumed to be normally-distributed, with a mean of zero (and a variance that can be estimated from the observations), and to be independent across observations and independent across values of the predictors (Bartholomew, Steele, Moustaki, & Galbraith, 2008).

Educational systems may unavoidably entail some similarities occurring between groups of students (Snijders & Bosker, 2012). For example, students within a school share the same environment, teachers, and geographical location. Accordingly, in ordinary-least-squares regression models, after controlling for the predictors, the values of the outcome may still be slightly dependent on the school (i.e. students within schools may have similar outcomes when compared across schools), hence residual errors may not be completely independent within schools (Snijders & Bosker, 2012). Generally, such similarities between students can entail lower standard errors associated with regression coefficients (the calculations of which involve residuals), which can entail increased chances for significance to be observed (i.e. lower *p*-values) (Bartholomew, Steele, Moustaki, & Galbraith, 2008; Snijders & Bosker, 2012).

Alternately, multi-level predictive modelling (also called hierarchical modelling or mixed modelling) can allow for similarities between students within groups such as schools (Hox, 2002; Snijders & Bosker, 2012). Essentially, residual errors are considered for schools and students per school: both are assumed to be normally-distributed and to have means of zero (and variances that can be estimated from the observations); however, these residuals can depend on the school (i.e. have separate distributions per school) (Snijders & Bosker, 2012). Residuals essentially represent unexplained variance, so a multi-level model can separate unexplained variance at the student-level from the school-level. Concurrently, multi-level models can account for further similarities or differences across schools. Specifically, schools can be modelled with varying regression constants (intercepts) and/or coefficients (slopes); practically, each varying term would be estimated as a 'random' variable with a specific distribution (rather than a 'fixed' parameter), from which schools may take different values (Snijders & Bosker, 2012).

Multi-level modelling was then selected for the final analysis to help account for residual similarities between students within schools. Given the focus on students' beliefs rather than school-level factors (such as resources available, entry requirements, etc.), the simplest multi-level models were applied (i.e. 'random-intercept' models, rather than 'random-slope' models, or models with 'random-intercepts and random-slopes'). The potential influence of school-level factors can be explored in various ways, such as through using aggregates of student-level factors or using discrete school-level indicators (Hox, 2002; Snijders & Bosker, 2012). However, considering school-level predictors would generally require large numbers of schools, which was not feasible for the new survey. Research in England has also suggested that schools' influences on choices may be more indirect than direct, and so can broadly be considered through students' views about potential areas, such as the provision of careers support (Bennett, Lubben, & Hampden-Thompson, 2013; Crawford, 2014; Gill & Bell, 2013). Fundamentally, and for consistency, the analysis focused on students' reported attitudes and beliefs (i.e. all the items/factors were consistently on the 'student' level), so that school-level factors were not considered.

Various other issues may still be relevant for predictive modelling, even within multi-level models. Simulations have highlighted that multi-level regression coefficients and their standard errors have been estimated without systemic bias, regardless of the number of groups (e.g. schools) considered, even for 5-10 groups; however, smaller numbers of groups (less than 50) can entail low standard errors associated with group-level variance components (Maas & Hox, 2005). Correlation between the various predictors is likely to be unavoidable in educational research, and extreme correlation between predictors has been referred to as 'multicollinearity'; while this does not violate any of the underlying assumptions, it can increase the standard errors of the regression coefficients (i.e. entailing higher *p*-values, and less likelihood of significant results) (Cohen & Cohen, 1984). Simulations have highlighted that multicollinearity has not appeared to bias parameter estimation within multi-level modelling (Yu, Jiang, & Land, 2015).

As a precaution, preliminary analysis considered predictive models through both single-level (via linear regression using ordinary-least-squares) and multi-level models; the parameters appeared broadly similar (e.g. excepting differences such as potentially more precise *p*-values in multi-level models), so that fundamental methodological issues did not appear to have been introduced. Residual plots were acceptable, and high multicollinearity was not present; 'tolerance' indicators, representing the proportion of variance for each predictor that was unexplained by the other predictors, were all above .1 (Cohen & Cohen, 1984).

Operationally, single-level regression involves an algebraic estimation process (i.e. ordinary-least-squares, where parameters can be calculated through formulae using observed features of the data) while multi-level regression involves iterative maximum-likelihood estimation in order to estimate the parameters that best fit the observed data (Snijders & Bosker, 2012). Various model indicators cannot be directly calculated via maximum-likelihood (e.g. adjusted R^2, which represents the proportion of variance in the outcome explained by the model). Accordingly, indicators of explained variance for the multi-level models were calculated as proportional reductions in residual variance (i.e. variance of the student-level and school-level residuals) when compared to a model without any predictors (Snijders & Bosker, 2012).

The predictive 'effect sizes' were calculated to reflect standardised coefficients (directly comparable across predictors): how many standard deviations of increase/decrease would occur in the outcome, given one standard deviation increase in the predictor (e.g. the estimated coefficient was multiplied by the standard deviation of the predictor, then divided by the standard deviation of the outcome) (Hox, 2002). Some studies have then multiplied the result by two (i.e. considering how many standard deviations of increase/decrease would occur in the outcome, given two standard deviation increases in the predictor), highlighting that it remains important to clarify or determine how effect sizes are defined (Tymms, 2004). Standardised coefficients are not effect sizes in the sense of the unique contributions/amounts of explained variance, although they may be broadly analogous to the correlation between two factors while controlling for one or more other factors. Accordingly, indicators of effect sizes for correlations were considered, specifically: above .1 as a 'small effect', above .3 as a 'medium effect', and above .5 as a 'large effect' (Cohen, 1988). Standardised coefficients of .1 and above were then considered to be meaningful.

In summary, multi-level modelling was selected for the final predictive models to help account for any residual similarities between students within schools. Preliminary analysis nevertheless applied (ordinary-least-squares) linear regression (confirming the inherent assumptions through various residual graphs) as a general sensitivity check, and the various parameters were sufficiently similar to suggest that no notable methodological issues were being introduced.

Section 6.3: Groups and clusters of students

Students can be categorised in various ways. Prior research has considered groups of under-confident, accurately-evaluating, and over-confident students, depending on group boundaries or criteria as defined by researchers (Dupeyrat, Escribe, Huet, & Régner, 2011; Gonida & Leondari, 2011). Alternately, cluster analysis allows students to be categorised according to their accuracy/bias without researchers needing to determine the specific cluster boundaries (Sáinz & Upadyaya, 2016). The proportions of students per cluster may vary, which may help determine how prevalent under-confidence or over-confidence may be.

For the PISA 2006 survey and for the 2014/2015 survey, students were accordingly categorised as conceptual ‘groups’ representing the conceptual/theoretical ideas of under-confidence, accuracy, and over-confidence (**Section 6.3.1**), and also as emergent ‘clusters’ of students from latent-profile analysis (**Section 6.3.2**).

Section 6.3.1: Conceptual groups

In the PISA 2006 survey, accuracy/bias was considered through a regression-residual indicator of self-concept accuracy/bias, as the questionnaire was not explicitly designed to measure accuracy/bias (**Section 5.2**). In the 2014/2015 survey, multiple indicators were available given the questionnaire design, and a difference-score indicator of task-confidence accuracy/bias was more contextually meaningful than a regression-residual indicator, and so was focused on for the final analysis (**Section 5.3**).

Conceptual groups were formed by directly categorising the accuracy/bias indicators, an approach that has been applied in various prior studies (Bouffard, Vezeau, Roy, & Lengelé, 2011; Gonida & Leondari, 2011; Narciss, Koerndle, & Dresel, 2011). The indicators were standardised as *z*-scores (per academic year, in the 2014/2015 survey) and ±.5 standard deviations were used as the group boundaries: below -.5 was classified as under-confident, between -.5 and +.5 was classified as accurate (one standard deviation range), and above +.5 was classified as over-confident. For greater insight, the accurately-evaluating students were divided into those with above-average task-scores and those with below-average task-scores (relative to the students' academic year in the 2014/2015 survey).

In PISA 2006, to provide a simple baseline for comparison (which also broadly reflected the profiles of the difference-score groups in the 2014/2015 survey), conceptual groups were also formed from cross-tabulating the students' self-concept and task-scores, when each was classified as either being above-average or below-average. Under-confidence was then defined as 'below-average self-concept with above-average task-scores', for example, while over-confidence was defined as 'above-average self-concept with below-average task-scores'. Accurately-low beliefs were considered as 'below-average self-concept with below-average task-scores', and accurately-high beliefs were considered as 'above-average self-concept with above-average task-scores'. This approach was broadly inspired by prior research that considered differences in above-average and below-average confidence and attainment via *z*-scores (Cheema & Skultety, 2016; Chiu & Klassen, 2010).

Additionally, preliminary analysis considered the implications of various different approaches to grouping, such as considering different group boundaries (i.e. ±1.0 standard deviations) and through grouping the various different indicators of accuracy/bias in the 2014/2015 survey in various ways. In the PISA 2006 survey, for example, the regression-residual group boundaries (±.5 standard deviations) produced relatively similar numbers of students per group which facilitated predictive modelling. Different boundaries (e.g. ±1.0 standard deviations) entailed broader definitions of 'accuracy' and hence classified far fewer students with extreme under-confidence or extreme over-confidence. Preliminary analysis highlighted that the patterns of predictive coefficients across the groups remained broadly similar regardless of different group boundaries (i.e. ±.5 or ±1.0).

In the 2014/2015 survey, cross-tabulations of the various conceptual accuracy/bias groups showed that the majority of students were similarly classified across approaches (grouping by difference-score or regression-residual) and when considering the task level (task-confidence and task-scores) or the subject level (self-concept beliefs and current grades), although there was still a notable amount of variation. Considering ±.5 standard deviations as the group boundaries highlighted that the majority of the accuracy/bias groups corresponded on the task and subject levels. The exception was that more 'over-confident' students from the subject-level self-concept/grade regression-residual groups were classified as 'accurate' students on the task-level difference-score groups. Considering ±1.0 standard deviations as the group boundaries highlighted that the majority of any task-level accuracy/bias group was classified as 'accurate' on the subject-level self-concept/grade regression-residual groups. Perhaps un-intuitively, these wider group boundaries may have entailed that accuracy was considered too broadly, rather than helping to reveal any common extremes of confidence biases. Preliminary analysis again highlighted that the patterns of predictive coefficients across the groups remained somewhat similar regardless of different group boundaries (i.e. ±.5 or ±1.0), although the results were likely less reliable due to the smaller numbers of extremely under-confident and extremely over-confident students when considering ±1.0 standard deviations as the group boundaries.

In summary, the final analysis of PISA 2006 focused on conceptual groups formed through the intersection of above-average/below-average self-concept and task-scores, and through the categorisation of the regression-residual indicator of self-concept accuracy/bias. The analysis of the 2014/2015 survey focused on the conceptual groups formed through the categorisation of the difference-score indicator of task-confidence accuracy/bias, which directly measured students' self-reflective accuracy/bias. The profile of above-average/below-average intersection groups in the PISA 2006 survey broadly reflected the profile of the difference-score groups in the 2014/2015 survey, which furthered comparability. These various groups provided exemplars of approaches from prior research that were most contextually relevant and meaningful given the two questionnaires designs. Preliminary analysis also explored various alternatives to grouping, but rather than undertake an extensive comparison of various (researcher-dependent) approaches to forming groups, cluster analysis was undertaken as the main alternate approach to categorising students.

Section 6.3.2: Emergent clusters

Cluster analysis allows students to be categorised according to their accuracy/bias through directly considering the underlying indicators of confidence and attainment, and avoiding the need to calculate and classify an indicator of accuracy/bias and the need for researchers to determine any specific group boundaries (Sáinz & Upadyaya, 2016).

Considered broadly, cluster analysis aims to identify naturally-occurring homogenous groups within data; for the specified criteria, clusters are found so that (essentially) those within clusters are similar to one another while any differences across clusters are maximised (Everitt, Landau, Leese, & Stahl, 2011). Cluster analysis can be accomplished by many different approaches, including algorithmic and model-based methods.

Algorithmic approaches include hierarchical clustering (e.g. via (agglomerative or divisive approaches) and optimisation clustering (e.g. using the *k*-means algorithm) (Everitt, Landau, Leese, & Stahl, 2011). Problematically, both hierarchical and optimisation clustering methods can be influenced by the order in which their algorithms encounter and then parse through the data (i.e. the cluster solution may slightly depend on the order of cases, which may be sorted by students/schools, by any item or factor, by random identifiers, etc.) (Everitt, Landau, Leese, & Stahl, 2011). Hierarchical cluster structures may be more appropriate for some contexts than others (e.g. biological classification, where species of animals can be hierarchically aggregated into clusters of genera, then families, orders, classes, phyla, etc.). Optimisation clustering involves specifying a number of clusters to be identified, and may be unhelpful without knowing plausible numbers of clusters to consider.

Model-based approaches to cluster analysis avoid these problems, and also provide indicators that can help determine the number of clusters that best fit the data (Collins & Lanza, 2010). Model-based cluster analysis has been called latent-class analysis when binary variables have been considered and latent-profile analysis when continuous or varying types of variables have been considered (Collins & Lanza, 2010; Masyn, 2013). Conceptually, model-based cluster analysis identifies clusters of students who have similar profiles of responses for the considered items/factors; ideally, homogenous clusters will be formed (i.e. within each cluster, the students have similar patterns of responses and hence there is little to no association between items/factors within the cluster) that are clearly separate to one another (i.e. different patterns or magnitudes of responses are characteristic of different clusters) (Collins & Lanza, 2010). Cluster membership then 'explains' the overall associations between factors seen in the entire sample: given the students' (conditional) cluster membership, the considered factors then have no association (are locally-independent) within the clusters themselves (i.e. there is an underlying assumption of 'conditional local independence') (Collins & Lanza, 2010). When considering multiple factors, those within a particular cluster may not necessarily have the same magnitude of responses for every factor; but they would have a particular profile of responses (e.g. which may involve high responses for some factors but low responses for others, or which may involve high response for all factors, or any particular profile given the data) that distinguishes them from the profiles of responses seen in other clusters.

Methodologically, cluster membership is modelled as a categorical latent-variable; iterative maximum-likelihood algorithms are used for model estimation; and classification of students to clusters is based on probabilities (i.e. each student has a varying probability, anywhere between 0 and 1, of belonging to each cluster) (Collins & Lanza, 2010). Model-based information criteria are usually used to compare models with different numbers of clusters, in order to consider which particular number of clusters may be best to explain the data (Collins & Lanza, 2010). Essentially, information criteria promote simplicity or parsimony (considered as 'estimating fewer parameters') when comparing different models. The Akaike Information Criterion (AIC) (Akaike, 1974) balances the fit of the model to the data and the complexity of the model (i.e. models with many parameters are 'penalised'); the Bayesian Information Criterion (BIC) (Schwarz, 1978) is similar, but penalises more parameters even more than the AIC. Simulation studies have resulted in recommendations to consider the BIC (Nylund, Asparouhov, & Muthén, 2007) or the sample-size adjusted BIC (Henson, Reise, & Kim, 2007). Nevertheless, and pragmatically, information criteria may be more useful in identifying plausible models rather than being able to definitively prove that any one single model is ideal (Collins & Lanza, 2010).

Various studies in educational and motivational research have applied hierarchical clustering (using Ward's method to help identify the optimal number of clusters to consider) (Ferguson & Bråten, 2013) and optimisation clustering (using the *k*-means algorithm) (Bøe & Henriksen, 2013). Other studies have applied model-based clustering such as latent-profile analysis (Chow, Eccles, & Salmela-Aro, 2012). Nevertheless, results from optimisation clustering (*k*-means) and model-based clustering (latent-class analysis) have been found to be broadly similar (DiStefano & Kamphaus, 2006).

For the analysis presented in this thesis, model-based clustering (i.e. latent-profile analysis) was selected in order to help identify optimal numbers of clusters via information criteria, and to avoid potential difficulties of algorithmic and optimisation clustering (e.g. results potentially varying depending on the order of cases in a data set). Latent-profile analysis had various other benefits, such as being able to potentially reveal a hierarchical structure or no specific structure of clusters, while algorithmic hierarchical clustering could only model/impose a hierarchical structure.

Latent-profile analysis was undertaken via Latent Gold software (Vermunt & Magidson, 2013). Sample-weighting was supported and applied when relevant (i.e. when analysing PISA 2006), but multiple-imputation was not supported (Vermunt & Magidson, 2013). However, even if multiple-imputation was supported to combine estimates of model parameters across the different plausible-values, selecting one particular plausible-value would still be necessary in order to export the classification of students to clusters, which appears to be the situation when using alternate software such as MPlus (Muthén & Muthén, 2012). Preliminary modelling of PISA 2006 therefore considered each plausible-value in turn (and together); given the similarity of the various results, the final analysis used the first plausible-value only.

Latent-profile analysis considered the students' self-concept beliefs and task-scores (first plausible-value) in the PISA 2006 survey, and considered their task-confidence and task-scores in the 2014/2015 survey (i.e. considering the same indicators of confidence and attainment used within the indicators of accuracy/bias). Operationally, the latent-profile analysis estimated various parameters: the mean per item/factor per cluster, and the error variance per item/factor per cluster, all while the covariance(s) between items/factors per cluster was set to zero (Vermunt & Magidson, 2013). This implemented the 'conditional local independence' assumption inherent to latent-profile analysis, although there still could be varying degrees of residual covariance between the considered items/factors. Various other models could be formed, but there appeared to be no prior precedent or theoretical reason to assume that they would be preferable (and the research did not attempt to undertake a technical exploration of different types of modelling). For example, 'conditional local independence' can be relaxed so that the covariance(s) between items/factors per cluster can be directly estimated rather than set to zero; this would technically be 'multivariate mixture' modelling rather than latent-profile analysis (Collins & Lanza, 2010; Vermunt & Magidson, 2013).

Since latent-profile analysis used maximum-likelihood estimation, it was possible that parameters could converge to different solutions depending on the starting values (i.e. potentially finding various 'locally-optimal' solutions rather than one single 'globally-optimal' solution) (Collins & Lanza, 2010). An extensive number of starting values (i.e. 1000) were therefore applied during estimation for each model; each set of models (i.e. with 1-10 clusters) was itself re-estimated 10 times to consider the consistency of the results. Consistency across these replications would not necessarily entail that a 'globally-optimal' solution was found (i.e. requiring an infinite number of replications), but could provide reassurance that solutions were sufficiently stable to be feasible and informative in practice. Only consistent models were considered within the final analysis; in practice, consistency appeared to only be an issue when modelling relatively high numbers of clusters.

In summary, cluster analysis was undertaken through latent-profile analysis in order to help identify optimal numbers of clusters to describe the data (via information criteria), and to avoid potential difficulties of algorithmic and optimisation clustering, such as results potentially varying depending on the order of cases in a data set, or needing to know the number of clusters to consider in advance (Collins & Lanza, 2010; Everitt, Landau, Leese, & Stahl, 2011).

Section 6.4: Aggregating academic years (2014/2015 survey)

The 2014/2015 survey was able to cover 1523 students (685 students in Year 9, 489 in Year 10, and 349 in Year 11). The power/sample size calculations underlying the research design (**Section 5.1**) indicated that considering more than 1152 students would be preferable for analysis, however. It was also less feasible to directly compare Year 11 students from PISA with the 349 Year 11 students from the 2014/2015 survey: dividing 349 students into groups would likely limit or prevent predictive modelling per group, given the low numbers involved.

Considered through analysis of variance, with Bonferroni post-hoc tests, the students in Years 10 and 11 in the 2014/2015 survey generally reported lower science attitudes, and science intentions, than students in Year 9. However, various patterns of differences were observed across the items/factors (tabulated in **Appendix 7** for reference). On average, students in Year 9 gained lower task-scores than students in Years 10 and 11, while students in Year 10 reported higher task-confidence than students in Years 9 and 11. On average, students in Years 10 and 11 exhibited small under-confidence biases while students in Year 9 were broadly accurate. Students in Year 10 had higher task sensitivity (accurately knowing when they had answered correctly) than students in Years 9 and 11. Students had similar task specificity (accurately knowing when they had answered incorrectly) across the different academic years. Students in Year 10 had higher task simple-matching (accurately knowing when they had answered correctly and incorrectly) than students in Year 9. Given the relatively small numbers of students involved (with different numbers per academic year), and since the data were not collected from the same students over time, conclusions could not easily be drawn regarding differences due to age.

Preliminary analysis highlighted that the various correlations between item/factors, and the patterns of coefficients when predicting science intentions, were similar when considered for each academic year separately. It was plausible to infer that students' intentions were relatively similarly influenced for each academic year. In order to maximise the number of considered students, the final analysis then considered students across Years 9, 10, and 11; the various accuracy/bias groups and clusters were formed relative to each academic year, but then considered together for all years. Essentially, the analysis explored differences in accuracy/bias via groups, but assumed that students' intentions were similarly predicted across any other potential groups (e.g. students' ages). Considering various other groups or moderation effects could form entire areas of analysis, such as considering whether science intentions were differently predicted across boys and girls, and could not be considered for reasons of brevity and since considering accuracy/bias groups by gender or other categories would entail insufficient group sizes for predictive modelling.

Within predictive models, the students' academic year was included in order to account for any varying magnitudes per item/factor per year. Essentially, the models then assumed that students' intentions would be similarly predicted by items/factors such as their interest across the different academic years, but accounted for the magnitudes of intentions, interest, and other items/factors potentially varying per academic year. Preliminary analysis confirmed that using per-year standardised predictors (i.e. the item/factors were converted to z-scores formed relative to each academic year) entailed the same conclusions as using the un-standardised predictors while also modelling the students' year as a predictor. Modelling the students' year (rather than standardising every item/factor per academic year) could also reveal any remaining influence due to students' ages.

The various accuracy/bias groups were formed through considering relative groups per year. For example, 'under-confidence' was defined relative to Year 9, 10, or 11, depending on the student, and then all the under-confident students in Years 9, 10, and 11 were considered together. This also helped account/control for any potential accuracy/bias differences linking with task ease/difficulty across the different years (**Section 3.2.1**). When considering differences in students' expressed attitudes and beliefs across the groups via analysis of variance tests, the indicators were standardised to reflect within-year differences (i.e. as *z*-scores formed per academic year). This could then consider whether the groups of students (across Years 9, 10, and 11) varied in their attitudes, as being above-average or below-average (in standard deviations) relative to the students' respective academic years, while accounting for the various means and standard deviations potentially varying per academic year.

Preliminary latent-profile analysis confirmed that the same clusters emerged per year regardless of whether indicators were standardised or not. Preliminary latent-profile analysis then considered standardised indicators (i.e. *z*-scores formed per academic year) across the separate academic years, and across multiple years considered together; broadly similar cluster profiles and proportions emerged. In order to maximise the number of considered students, the final latent-profile analysis then considered standardised indicators (i.e. *z*-scores formed per academic year) and students across multiple academic years (i.e. Years 9, 10, and 11).

In summary, for the analysis of the 2014/2015 survey, students were considered across Years 9, 10, and 11, in order to ensure that sufficient numbers of students were considered to reveal potential effects. The groups and clusters were formed to represent within-year differences, and then considered together across these three academic years, which accounted/controlled for any differences in means/responses across the years.

Section 7: Research question one

The first research question considered which attitudes and motivational beliefs, including different conceptualisations of confidence and indicators of confidence accuracy/bias, were the most relevant influences on students' intentions and choices. The results from the PISA 2006 survey (**Section 7.1**) and the 2014/2015 survey (**Section 7.2**) are described in turn, and then discussed and contextualised in **Section 7.3**.

When applied to predictive modelling, the earlier conceptual model (**Figure 1**) broadly supported a series of models, starting with modelling the predictive associations between students' background characteristics and their science intentions, and then sequentially including various further factors as additional predictors (specifically, confidence and attainment, theorised influences ('subjective task values') including interest and utility value, and then further potential influences). Any reductions in the predictive coefficients from one step to the next would suggest that the original associations were 'mediated' (or explained) by differences in the new factors (Baron & Kenny, 1986; MacKinnon, Kisbu-Sakarya, & Gottschall, 2013). However, given the overall analytical approach to addressing the first research question (i.e. ultimately determining which factors had the highest predictive associations with science intentions; **Section 6**), the predictive models did not attempt to consider every potential association within the conceptual model (**Figure 1**); any mediation associations would be implicit, observed through any varying coefficients at different stages, rather than explicitly considered through also modelling predictors of students' confidence, attainment, and attitudes.

Sequential modelling nevertheless aimed to provide insight at each step. Specifically, initial modelling determined the predictive associations between students' background characteristics and intentions (determining which factors were most predictive and to what extent, while accounting for the others); this also facilitated contextualisation of the modelled students against any other research that may have only considered particular aspects of students' background (such as gender). Subsequently, also including students' confidence and attainment within the model then considered their potential motivational implications, with and without accounting for the other (which explored whether confidence appeared motivational when considered alone through partially reflecting attainment, and/or vice versa). Subsequently, also including further predictors (such as interest in science and perceived utility of science) then considered whether the apparent motivational effect of confidence was reduced (where the previous effect of confidence may have partially encompassed or reflected another factor such as interest value or utility value, given potentially-close associations between students' attitudes and beliefs), while also determining the independent predictive associations of the further factors (especially interest value, utility value, and other potentially-relevant predictors from prior research; **Section 2**). Subsequently, any further factors were then also included in order to provide a comprehensive model to determine which factors ultimately had the highest predictive associations with science intentions.

Section 7.1: Results: PISA 2006

The (linear) associations between students' science intentions and attitudes can be described through (Pearson) correlation coefficients. The highest correlations with students' science intentions occurred with science utility value (R = .684, p < .001), science interest/enjoyment value (R = .610, p < .001), personal value of science (R = .607, p < .001), science self-concept beliefs (R = .537, p < .001), and interest across the various science areas/topics (R = .516, p < .001). Students' self-concept beliefs also most strongly correlated with their interest/enjoyment value (R = .620, p < .001), personal value (R = .555, p < .001), self-efficacy for various types of science tasks/activities (R = .534, p < .001), and utility value (R = .531, p < .001). **Table 4** summarises the various correlations for the main measures of attitudes and beliefs (limited for brevity).

Students' attitudes and beliefs may closely associate in complex ways, so that predictive modelling was necessary in order to isolate the independent associations of the various factors. For example, students' interest directly correlated with their intentions, but higher interest may also foster higher self-concept beliefs (or vice versa) and so also indirectly associate with intentions, given that interest correlated with self-concept beliefs and that self-concept beliefs correlated with intentions.

For greater insight and contextualisation, a series of predictive models were then considered (**Table 5**). Within these models, the students' reported science intentions, represented as a continuous/scalar factor-score (**Section 5.2**), were predicted using multi-level linear (predictive) modelling (**Section 6.2**). Through the use of expectation-maximisation to estimate missing responses (**Section 6.1**), the models were able to consistently consider 4645 of the 4935 students surveyed in England in PISA 2006, regardless of the inclusion of different predictors at different steps within the process; unavoidably, missing binary indicators such as gender could not be estimated, and so the number of considered students was still slightly reduced. However, without estimating missing responses, only 3860 students would have been considered in the final model, and different numbers of students could have been considered at different steps depending on which predictors were included.

When predicting science intentions for all students, background characteristics such as their gender, whether their parents worked within science or not, and the number of books at home only explained a small amount of variance (**Table 5**, step 1). Including the students' (plausible-value) task-score (but not the students' self-concept beliefs) only provided a modest improvement in explained variance (**Table 5**, step 2a). However, the students' attainment appeared to 'mediate' the influences of the number of books at home and parents working within science (Baron & Kenny, 1986; MacKinnon, Kisbu-Sakarya, & Gottschall, 2013); essentially, these factors could be interpreted to associate with attainment (task-scores), which then associated with intentions (although any causal relations cannot be confirmed since the factors were all measured at the same time). Alternately, including the students' self-concept beliefs (but not the students' task-score) explained more variance (**Table 5**, step 2b), but only mediated the association between intentions and gender; essentially, the association between gender and intentions (i.e. boys were predictively associated with higher intentions at step 1) appeared to follow from the underlying association between intentions and self-concept beliefs (i.e. where boys reported higher self-concept beliefs).

Including the students' task-score and self-concept beliefs (**Table 5**, step 3) highlighted that self-concept beliefs had a relatively higher predictive association with intentions than task-scores, and highlighted little difference in the proportion of explained variance compared to modelling self-concept beliefs without task-scores (i.e. step 3 compared to step 2b). This potentially highlighted the motivational nature of self-concept beliefs, and hence the potential importance of considering the impact of under-confidence or over-confidence.

Including the available factors from the expectancy-value model allowed over half of the variance in the students' intentions to be explained (**Table 5**, step 4). These factors also appeared to mediate the association between students' intentions and self-concept beliefs (i.e. the predictive coefficient for self-concept was lower in step 4 compared to step 3, but still significant); essentially, higher self-concept beliefs may partially reflect higher interest, utility, and/or personal value. Additional models (not tabulated for brevity) highlighted that interest, utility, and personal value each mediated the predictive association between self-concept beliefs (and for some other factors) and intentions, and utility value appeared to entail the largest change (i.e. the reduction of the predictive coefficient between self-concept beliefs and intentions). Including further potential influences such as the frequency of various teaching approaches produced little change in the proportion of explained variance (**Table 5**, step 5).

Further analysis would be necessarily to isolate and determine mediation relations in more detail, but such areas were outside the scope of the immediate research questions. Essentially, the sequential models provided context through helping to plausibly explain why aspects of students' background (e.g. their gender and whether their parents worked within science) were not directly predictive of intentions in the final model. The models also highlighted that self-concept beliefs may reflect interest, utility, and other factors that have not been considered as theoretical antecedents to expressions of confidence, as considered in more detail elsewhere (Sheldrake, 2016a, 2016b). On a wider level, the results did not appear to be sensitive to the analytical approaches and methodology; preliminary analysis considered various alternate approaches, for example using multiple-imputation across all task-score plausible-values and modelling without estimating missing responses, which still produced the same results as reported here.

Fundamentally, the first research question was addressed by the final predictive model (**Table 5**, step 5), where students' science intentions were most strongly predicted by their perceived utility of science, interest/enjoyment of science, personal value of science, and their science self-concept beliefs. Various other items/factors were predictive, but at lower relative magnitudes (a standardised predictive coefficient of .1 was considered the threshold for a meaningful magnitude; **Section 6.2**). Ultimately, students' background and gender were not predictive of their science intentions when accounting for their attitudes and motivational beliefs, which cohered with findings from prior studies (DeWitt, Archer, & Osborne, 2014; Mujtaba & Reiss, 2014). Similarly, the pattern of predictors, broadly highlighting the importance of students' perceived utility, interest, and (self-concept) confidence, cohered with prior findings (Mujtaba & Reiss, 2014; Wang & Degol, 2013). The low predictive association between the students' task-scores and science intentions may potentially highlight that the non-curricular nature of the OECD tasks limits their contextual relevance. Students' examination grades, for example, might show higher associations with their intentions, given that current or expected grades may directly facilitate or preclude students' options (Brown, Brown, & Bibby, 2008; Department for Children, Schools and Families, 2009). While such results suggested that any subsequent analytical models could be greatly simplified to only consider the strongest identified predictors, it was possible that other factors were more relevant to specific groups of students when considered alone.

The regression-residual indicator of accuracy/bias could not be modelled as a predictor of science intentions together with the students' self-concept beliefs and task-scores. Any one of the indicators of accuracy/bias, self-concept, and task-score could be perfectly predicted given the other two indicators, which entailed the highest possible 'multicollinearity' within a predictive model (Cohen & Cohen, 1984). This situation would be identified automatically, and one of the three indicators would be removed by statistical software if all three were included as predictors.

It was still possible to model the various combinations of self-concept, task-scores, and the accuracy/bias indicator, considering any two of the three factors (**Table 6**). Specifically, the various models highlighted that, controlling for the various other predictors: task-scores and self-concept beliefs both positively predicted science intentions (**Table 6**, model A); task-scores and the self-concept accuracy/bias indicator both positively predicted science intentions (i.e. higher over-confidence, controlling for task-scores, predicted higher science intentions; **Table 6**, model B); self-concept positively predicted and the accuracy/bias indicator negatively predicted science intentions (i.e. higher over-confidence, controlling for self-concept beliefs, predicted lower science intentions; **Table 6**, model C). These patterns were not necessarily easy to interpret, however, hence the impact of under-confidence and over-confidence on students' intentions needed to be considered through the other research questions via considering groups and clusters of students.

Table 4: PISA (England) 2006: items/factors, selected correlations

Factor/scale	1	2	3	4	5	6	7	8	9	10
1. Gender (1=boy)	1.000									
2. Intentions	.092	1.000								
3. Task-score (PV1)	.051	.260	1.000							
4. Self-concept	.203	.537	.331	1.000						
5. Self-concept accuracy/bias	.197	.473	(<.001)	.943	1.000					
6. Self-efficacy (areas)	.132	.345	.528	.534	.375	1.000				
7. Interest (areas)	.071	.516	.242	.484	.424	.413	1.000			
8. Interest value	.126	.610	.366	.620	.525	.462	.650	1.000		
9. Utility value	.067	.684	.227	.531	.477	.333	.515	.571	1.000	
10. Personal value	.091	.607	.300	.555	.477	.467	.546	.630	.645	1.000
11. General value	.129	.358	.362	.417	.310	.465	.397	.476	.406	.655

Notes: Missing responses were estimated by expectation-maximisation. Pearson's correlation coefficients (R) are reported. All coefficients were significant ($p < .001$) except when highlighted in brackets (i.e. for task-score and the self-concept accuracy/bias indicator: $R < .001$, $p = 1.000$).

Table 5: PISA (England) 2006: predicting science intentions

Item/factor	Step 1 Std. Effect	Step 1 Sig. (*p*)	Step 2a Std. Effect	Step 2a Sig. (*p*)	Step 2b Std. Effect	Step 2b Sig. (*p*)	Step 3 Std. Effect	Step 3 Sig. (*p*)	Step 4 Std. Effect	Step 4 Sig. (*p*)	Step 5 Std. Effect	Step 5 Sig. (*p*)
Intercept/constant	**NA**	**<.001**	**NA**	**<.001**	**NA**	**<.001**	**NA**	**<.001**	**NA**	**<.001**	**NA**	**<.001**
Gender (1=boy)	**.100**	**<.001**	**.084**	**<.001**	-.016	.220	-.016	.222	-.002	.843	.006	.543
Books at home	**.089**	**<.001**	-.010	.536	**.031**	**.024**	-.002	.904	-.010	.387	-.016	.155
Parent(s) in science (1=yes)	**.042**	**.005**	.017	.258	**.026**	**.043**	.018	.157	.006	.590	.006	.550
Mothers' education	-.033	.058	-.030	.069	**-.030**	**.039**	**-.030**	**.038**	-.004	.734	.001	.958
Fathers' education	**.109**	**<.001**	**.097**	**<.001**	**.061**	**<.001**	**.059**	**<.001**	.014	.239	.008	.494
Task-score (PV1)			**.280**	**<.001**			**.095**	**<.001**	.018	.129	**.061**	**<.001**
Self-concept					**.533**	**<.001**	**.506**	**<.001**	**.101**	**<.001**	**.111**	**<.001**
Self-efficacy (areas)											**-.050**	**<.001**
Interest (various areas)											**.055**	**<.001**
Interest value									**.220**	**<.001**	**.189**	**<.001**
Utility value									**.405**	**<.001**	**.393**	**<.001**
Personal value									**.147**	**<.001**	**.182**	**<.001**
General value											**-.087**	**<.001**
Science activities											**.069**	**<.001**
School career preparation											**-.030**	**.014**
School career information											**.085**	**<.001**
Teaching: interaction											**-.037**	**.004**
Teaching: activities											**-.042**	**.001**
Teaching: investigations											.022	.067
Teaching: applications											**-.044**	**.001**
Explained variance	3.1%		8.2%		29.9%		30.5%		56.2%		58.4%	
Unexplained variance (residual)	94.0%		88.0%		68.9%		68.2%		43.6%		41.4%	
Unexplained variance (school)	2.8%		3.8%		1.2%		1.4%		.2%		.2%	

Notes: Missing responses were estimated by expectation-maximisation. Standardised coefficients ('Std. Effect') and significance (p-values; 'Sig. (p)') are reported. Significant coefficients (p < .05) are highlighted in bold.

Table 6: PISA (England) 2006: predicting science intentions, including the accuracy/bias indicator

	Model A		Model B		Model C	
Item/factor	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)
Intercept/constant	**NA**	**<.001**	**NA**	**<.001**	NA	.001
Gender (1=boy)	.006	.543	.005	.597	.005	.597
Books at home	-.016	.155	-.015	.180	-.015	.180
Parent(s) in science (1=yes)	.006	.550	.003	.751	.003	.751
Mothers' education	.001	.958	.005	.691	.005	.691
Fathers' education	.008	.494	.006	.597	.006	.597
Task-score (PV1)	**.061**	**<.001**	**.101**	**<.001**		
Self-concept	**.111**	**<.001**			**.294**	**<.001**
Self-concept accuracy/bias			**.104**	**<.001**	**-.177**	**<.001**
Self-efficacy (areas)	**-.050**	**<.001**	**-.053**	**<.001**	**-.053**	**<.001**
Interest (various areas)	**.055**	**<.001**	**.059**	**<.001**	**.059**	**<.001**
Interest value	**.189**	**<.001**	**.185**	**<.001**	**.185**	**<.001**
Utility value	**.393**	**<.001**	**.390**	**<.001**	**.390**	**<.001**
Personal value	**.182**	**<.001**	**.188**	**<.001**	**.188**	**<.001**
General value	**-.087**	**<.001**	**-.088**	**<.001**	**-.088**	**<.001**
Science activities	**.069**	**<.001**	**.069**	**<.001**	**.069**	**<.001**
School career preparation	**-.030**	**.014**	**-.027**	**.027**	**-.027**	**.027**
School career information	**.085**	**<.001**	**.083**	**<.001**	**.083**	**<.001**
Teaching: interaction	**-.037**	**.004**	**-.037**	**.005**	**-.037**	**.005**
Teaching: activities	**-.042**	**.001**	**-.042**	**.001**	**-.042**	**.001**
Teaching: investigations	.022	.067	.024	.055	.024	.055
Teaching: applications	**-.044**	**.001**	**-.044**	**.002**	**-.044**	**.002**
Explained variance	58.4%		58.3%		58.3%	
Unexplained variance (residual)	41.4%		41.5%		41.5%	
Unexplained variance (school)	.2%		.2%		.2%	

Notes: Missing responses were estimated by expectation-maximisation. Standardised coefficients ('Std. Effect') and significance (p-values; 'Sig. (p)') are reported. Significant coefficients (p < .05) are highlighted in bold.

Section 7.2: Results: 2014/2015 survey

The 2014/2015 survey was analysed similarly to the PISA 2006 survey, through first considering correlations and then predictive modelling.

Considered across Years 9-11, the strongest correlations with students' science intentions were observed for the students' perceived utility of science (R = .732, p < .001), personal value of science (R = .684, p < .001), interest value of science (R = .622, p < .001), and the subjective-norm/influence of parents (R = .554, p < .001). Moderate correlations were also observed between intentions and students' self-concept (R = .475, p < .001) and self-efficacy beliefs (R = .444, p < .001), and also between intentions and self-regulated studying (R = .492, p < .001), and the strategies of elaboration (R = .448, p < .001), control (R = .444, p < .001), and memorisation (R = .435, p < .001). A slightly lower association was observed between intentions and students' current science grade (R = .351, p < .001). Intentions also had a higher association with task-confidence (R = .403, p < .001) than task-score (R = .256, p < .001).

The higher associations between science intentions and different indicators of 'confidence' (i.e. self-concept, self-efficacy, and task-confidence) compared to 'attainment' (i.e. current grades and task-scores) perhaps highlighted the potential relevance of confidence accuracy/bias. The difference-score indicator of task accuracy/bias (i.e. under-confidence through accuracy through to over-confidence) nevertheless had a minimal association with science intentions (R = .080, p = .002). However, task sensitivity (R = .259, p < .001), specificity (R = -.317, p < .001), and simple-matching (R = .115, p < .001) showed relatively higher associations with intentions. Comprehensive correlation tables are summarised in **Appendix 8**.

As with PISA 2006, predictive models were applied in order to isolate the independent associations between intentions and students' various attitudes and beliefs. Initial comparison of equivalent models (**Table 7**) highlighted that the samples of students differed across the two surveys: specifically, self-concept beliefs and interest value appeared to be relatively less predictive of science intentions in the 2014/2015 survey than in the PISA 2006 survey, while personal value appeared to be relatively more predictive. Accordingly, some differences in results were expected across the two surveys. Limiting the analysis to only consider equivalent factors across the two surveys would have undermined the purpose of directly measuring self-reflective accuracy/bias through task-confidence and task-scores, and further potential influences on intentions, via the 2014/2015 survey.

A series of linear multi-level predictive models were again applied for the 2014/2015 survey, to increase insight and provide wider contextualisation (**Table 8**). As with the PISA 2006 survey, expectation-maximisation was applied to estimate missing responses, which allowed 1423 (rather than only 571) of the 1523 students to be considered, regardless of which predictors were included.

When predicting science intentions, students' background characteristics only explained a small amount of variance (**Table 8**, step 1). Extending from this step (not tabulated for brevity), students' expressions of confidence (task-confidence, self-concept beliefs, self-efficacy beliefs) and indicators of attainment (task-scores, current science grades) were each positively predictive of intentions, accounting for the students' background characteristics, when considered in turn. As with PISA 2006, the indicators of attainment each appeared to mediate the influence of the number of books at home, while the indicators of confidence each appeared to mediate the influence of gender. The students' current grade was positively predictive when modelled with the students' background and their self-concept beliefs or with the students' background and their self-efficacy beliefs. However, current grades lost significance when modelled with the students' background, self-concept beliefs, and self-efficacy beliefs.

Modelling the various indicators of students' background, confidence, and attainment (**Table 8**, step 2) highlighted that self-concept and self-efficacy had the highest predictive associations with intentions. Adding the theorised influences from the expectancy-value model (**Table 8**, step 3), specifically interest value, utility value, personal value, and (the absence of) costs, mediated the influences of parents working in science, self-concept beliefs, and (to some extent) self-efficacy beliefs. As with PISA 2006, expressions of confidence may partially reflect interest and other factors, which may help explain why they appear motivational.

Adding the theorised influences in turn instead of together (not tabulated for brevity) highlighted that: utility value appeared to be the main mediator of parents working in science; additionally, interest, utility, and personal value each mediated the predictive association between self-concept beliefs and intentions, with interest and then utility entailing the largest changes.

The various factors from the expectancy-value model (**Table 8**, step 3) explained more variance than other potential predictors (**Table 8**, step 4), such as focusing on mastering science work, believing that effort can achieve success in science (perceived control), and the various studying strategies. Nevertheless, some of these predictors were still significant when considered together with the expectancy-value factors (**Table 8**, step 5), although with relatively low predictive magnitudes.

Finally, when considering the comprehensive set of items/factors (**Table 8**, step 6), students' science intentions were most strongly predicted by the students' perceived utility of science, personal value of science, self-efficacy, subjective-norms/influences from parents, and the students' interest value (which was only just below the .1 coefficient magnitude threshold). Various other items/factors were predictive, but at lower relative magnitudes. Despite the various study strategies having moderate correlations with science intentions, only the memorisation studying strategy had a minimal predictive association within the final model. In contrast to self-efficacy, neither the students' self-concept beliefs nor their current science grades were predictive of their science intentions when accounting for the various other factors. An alternate model highlighted that self-concept beliefs (but not the students' current grade) was predictive if self-efficacy was removed (standardised coefficient for self-concept = .058, p = .032, at the equivalent of step 6, but without self-efficacy; the other predictors were essentially unchanged). Given that self-concept beliefs have been predictive of science intentions in prior research in England (DeWitt, Archer, & Osborne, 2014; Mujtaba & Reiss, 2014), and in the PISA 2006 data, the different implication of self-concept, self-efficacy, and students' grades likely need to be explored in more detail with further students.

As before with PISA 2006, the task-confidence (difference-score) accuracy/bias indicator could not be modelled together with the task-score and task-confidence indicators. When considering the comprehensive set of predictors (i.e. step 6), but including various combinations of task-score, task-confidence, and (difference-score) accuracy/bias (not tabulated for brevity): task-score was positively predictive of intentions while task-confidence was not significantly predictive; replacing task-score and task-confidence with the task accuracy/bias indicator highlighted that it was not significantly predictive; when included together, task-score was positively predictive of intentions but task accuracy/bias was not significantly predictive; finally, when included together, task-confidence was positively predictive while task accuracy/bias was negatively predictive. The various coefficient magnitudes were very small, however. The same pattern was seen when using the task-confidence regression-residual accuracy/bias indicator. However, no elements of the equivalent combinations of self-concept, current grades, and self-concept (regression-residual) accuracy/bias were significant.

For contextualisation (**Table 9**), when modelled alone with the students' background characteristics, the various patterns of coefficients for the task-level indicators were more clearly revealed. In addition to the students' background characteristics, task-score and task-confidence both positively predicted science intentions (**Table 9**, model A); task-scores and the (difference-score) accuracy/bias indicator both positively predicted science intentions (i.e. higher over-confidence, controlling for task-scores, predicted higher science intentions; **Table 9**, model B); task-confidence positively predicted science intentions and the accuracy/bias indicator negatively predicted science intentions (i.e. higher over-confidence, controlling for task-confidence, predicted lower science intentions; **Table 9**, model C). This also mirrored the equivalent patterns in PISA 2006 (**Table 6**). Similarly, when modelled with students' background characteristics, the indicators of task sensitivity, task specificity, and task simple-matching, and also the regression-residual accuracy/bias indicators, were predictive of science intentions (**Table 10**). However, they were not significantly predictive when the students' other attitudes and beliefs were also included as predictors (not tabulated for brevity).

Predictive models and results cannot be selected simply to provide significant results, however, and it was possible to infer that indicators of self-reflective accuracy/bias on the task-level did not necessarily have a large direct influence on students’ subject-level science studying intentions. The analysis continued, nevertheless, to consider whether task-level self-reflective confidence accuracy/bias had an indirect or other influence on students’ intentions: groups or clusters of students with different degrees of self-reflective accuracy/bias might be more suited to explore the area (covered by research question two, **Section 8**), and tendencies towards under-confidence or over-confidence might entail that students’ intentions were predicted in different ways (covered by research question three, **Section 9**).

Table 7: England 2014/2015 (Years 9, 10, 11): predicting science intentions (contextualisation with PISA 2006)

	PISA 2006 survey		2014/2015 survey	
Item/factor	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)
Intercept	**NA**	**<.001**	NA	.404
Year	NA	NA	-.007	.715
Gender (1=boy)	-.002	.846	.022	.248
Books at home	-.004	.726	-.020	.294
Parent(s) in science (1=yes)	.007	.497	.027	.127
Mothers' education	-.004	.751	-.006	.797
Fathers' education	.014	.225	.012	.614
Self-concept	**.104**	**<.001**	**.059**	**.010**
Interest value	**.224**	**<.001**	**.072**	**.016**
Utility value	**.404**	**<.001**	**.467**	**<.001**
Personal value	**.148**	**<.001**	**.221**	**<.001**
Explained variance	56.2%		58.9%	
Unexplained variance (residual)	43.6%		40.5%	
Unexplained variance (school)	.2%		.6%	

Notes: Missing responses were estimated by expectation-maximisation. Standardised coefficients ('Std. Effect') and significance (p-values; 'Sig. (p)') are reported. Significant coefficients (p < .05) are highlighted in bold.

Table 8: England 2014/2015 (Years 9, 10, 11): predicting science intentions

	Step 1		Step 2		Step 3		Step 4		Step 5		Step 6	
Item/factor	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)
Intercept	**NA**	**<.001**	**NA**	**.013**	**NA**	**.007**	NA	.487	NA	.150	NA	.110
Year	**-.088**	**.001**	-.045	.076	.011	.571	**-.074**	**.001**	.003	.887	.009	.637
Gender (1=boy)	**.091**	**.001**	-.002	.920	.011	.543	.039	.102	.010	.605	.002	.915
Ethnicity (Black)	.026	.336	.023	.326	.026	.142	.013	.561	.020	.239	.019	.253
Ethnicity (East-Asian)	.025	.336	.012	.601	.001	.959	.008	.715	-.004	.797	-.008	.636
Ethnicity (South-Asian/Indian)	**.197**	**<.001**	**.132**	**<.001**	**.075**	**.001**	**.120**	**<.001**	**.057**	**.004**	**.049**	**.013**
Ethnicity (mixed)	.038	.141	.024	.283	.030	.085	.032	.140	.024	.166	.021	.218
Ethnicity (other)	**.064**	**.011**	**.048**	**.030**	**.040**	**.016**	**.053**	**.013**	.032	.053	.029	.069
Books at home	**.074**	**.009**	-.035	.180	-.037	.058	.005	.823	**-.044**	**.024**	**-.042**	**.028**
Parent(s) in science (1=yes)	**.065**	**.013**	**.057**	**.013**	.026	.139	**.051**	**.022**	.026	.128	.014	.409
Mothers' education	-.030	.390	**-.069**	**.027**	-.016	.492	-.039	.188	-.018	.452	-.017	.452
Fathers' education	**.095**	**.010**	.020	.541	-.003	.908	.023	.458	.001	.980	-.013	.580
Task-score			.056	.055	**.047**	**.034**	.049	.070	**.065**	**.003**	**.058**	**.007**
Task-confidence			**.119**	**<.001**	.013	.596	**.112**	**<.001**	.024	.324	.022	.361
Current grade			-.020	.569	-.029	.282			-.036	.162	-.037	.159
Self-concept			**.280**	**<.001**	.012	.639			.033	.202	.037	.173
Self-efficacy			**.203**	**<.001**	**.110**	**<.001**			**.135**	**<.001**	**.118**	**<.001**
Interest value					.049	.102			**.087**	**.006**	**.097**	**.003**
Utility value					**.463**	**<.001**			**.508**	**<.001**	**.451**	**<.001**
Personal value					**.231**	**<.001**			**.211**	**<.001**	**.215**	**<.001**
Cost value (absence of)					**.038**	**.041**			**.059**	**.002**	**.052**	**.007**
Orientation: mastery							**.083**	**.002**	**-.053**	**.015**	-.040	.068
Orientation: performance							.049	.057	**-.054**	**.008**	**-.053**	**.010**
Perceived control							**.077**	**.004**	**-.073**	**.001**	**-.050**	**.029**
Perceived control (exams)							.023	.339	**-.056**	**.006**	**-.059**	**.005**
Study strategy: self-regulation							**.147**	**<.001**	-.013	.672	-.005	.875

Item/factor	Step 1 Std. Effect	Step 1 Sig. (*p*)	Step 2 Std. Effect	Step 2 Sig. (*p*)	Step 3 Std. Effect	Step 3 Sig. (*p*)	Step 4 Std. Effect	Step 4 Sig. (*p*)	Step 5 Std. Effect	Step 5 Sig. (*p*)	Step 6 Std. Effect	Step 6 Sig. (*p*)
Study strategy: control							.040	.272	.006	.821	.036	.205
Study strategy: memorisation							.066	.056	**.078**	**.004**	**.073**	**.007**
Study strategy: elaboration							**.117**	**<.001**	-.012	.634	-.018	.483
Anxiety (absence of)											.038	.164
Mastery norms (good grade)											.005	.766
Subject-comparisons											.045	.060
Peer-comparisons											-.012	.617
Social persuasions (praise)											**-.052**	**.033**
Vicarious experiences											**-.063**	**.002**
Norms/influence (friends)											**-.049**	**.009**
Norms/influence (parents)											**.109**	**<.001**
Teacher perceptions											-.046	.088
Teacher/school careers/events											**.046**	**.039**
Explained variance	10.9%		31.4%		60.6%		36.3%		61.8%		63.2%	
Unexplained variance (residual)	86.2%		68.4%		39.3%		63.7%		38.2%		36.8%	
Unexplained variance (school)	2.9%		.3%		.0%		.0%		.0%		.0%	

Notes: Missing responses were estimated by expectation-maximisation. Standardised coefficients ('Std. Effect') and significance (p-values; 'Sig. (p)') are reported. Significant coefficients (p < .05) are highlighted in bold.

Table 9: England 2014/2015 (Years 9, 10, 11): predicting science intentions, considering students' background and task accuracy/bias

	Model A		Model B		Model C	
Item/factor	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)
Intercept	**NA**	**<.001**	**NA**	**<.001**	**NA**	**<.001**
Year	**-.128**	**<.001**	**-.128**	**<.001**	**-.128**	**<.001**
Gender (1=boy)	.015	.574	.015	.574	.015	.574
Ethnicity (Black)	.024	.340	.024	.340	.024	.340
Ethnicity (East-Asian)	.001	.958	.001	.958	.001	.958
Ethnicity (South-Asian/Indian)	**.163**	**<.001**	**.163**	**<.001**	**.163**	**<.001**
Ethnicity (mixed)	.039	.111	.039	.111	.039	.111
Ethnicity (other)	**.057**	**.016**	**.057**	**.016**	**.057**	**.016**
Books at home	.004	.892	.004	.892	.004	.892
Parent(s) in science (1=yes)	**.056**	**.024**	**.056**	**.024**	**.056**	**.024**
Mothers' education	-.052	.112	-.052	.112	-.052	.112
Fathers' education	.054	.119	.054	.119	.054	.119
Task-score	**.104**	**.001**	**.495**	**<.001**		
Task-confidence	**.319**	**<.001**			**.403**	**<.001**
Task-confidence accuracy/bias (difference-score)			**.354**	**<.001**	**-.094**	**.001**
Explained variance	22.2%		22.2%		22.2%	
Unexplained variance (residual)	77.1%		77.1%		77.1%	
Unexplained variance (school)	.7%		.7%		.7%	

Notes: Missing responses were estimated by expectation-maximisation. Standardised coefficients ('Std. Effect') and significance (p-values; 'Sig. (p)') are reported. Significant coefficients (p < .05) are highlighted in bold.

Table 10: England 2014/2015 (Years 9, 10, 11): predicting science intentions, considering students' background and further accuracy/bias indicators

	Model A		Model B		Model C		Model D		Model E		Model F		Model G	
Item/factor	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)
Intercept	**NA**	**<.001**	**NA**	**<.001**	**NA**	**<.001**	**NA**	**<.001**	**NA**	**<.001**	**NA**	**<.001**	**NA**	**<.001**
Year	**-.128**	**<.001**	**-.083**	**.002**	**-.106**	**<.001**	**-.088**	**<.001**	**-.096**	**<.001**	**-.100**	**<.001**	**-.097**	**.001**
Gender (1=boy)	.015	.574	**.082**	**.004**	.048	.068	.040	.130	**.085**	**.002**	.039	.149	.051	.093
Ethnicity (Black)	.024	.340	.026	.335	.025	.318	.024	.334	.026	.334	.024	.359	.015	.621
Ethnicity (East-Asian)	.001	.958	.025	.345	.018	.482	.015	.536	.025	.341	.017	.511	.051	.070
Ethnicity (South-Asian/Indian)	**.163**	**<.001**	**.196**	**<.001**	**.192**	**<.001**	**.179**	**<.001**	**.198**	**<.001**	**.188**	**<.001**	**.164**	**<.001**
Ethnicity (mixed)	.039	.111	.037	.147	.039	.118	.033	.175	.039	.134	.036	.153	.027	.309
Ethnicity (other)	**.057**	**.016**	**.064**	**.011**	**.059**	**.015**	**.064**	**.007**	**.064**	**.010**	**.061**	**.011**	.058	.034
Books at home	.004	.892	**.079**	**.006**	**.055**	**.046**	**.072**	**.008**	**.072**	**.010**	**.075**	**.006**	.048	.111
Parent(s) in science (1=yes)	**.056**	**.024**	**.064**	**.014**	**.059**	**.022**	**.064**	**.011**	**.063**	**.016**	**.062**	**.016**	**.098**	**<.001**
Mothers' education	-.052	.112	-.029	.400	-.045	.192	-.025	.449	-.033	.344	-.031	.372	-.064	.078
Fathers' education	.054	.119	**.089**	**.015**	**.088**	**.014**	**.077**	**.029**	**.091**	**.013**	.063	.078	.057	.134
Task-score	**.104**	**.001**												
Task-confidence	**.319**	**<.001**												
Task-conf. accuracy/bias (diff.-score)			**.054**	**.038**										
Task sensitivity					**.189**	**<.001**								
Task specificity							**-.255**	**<.001**						
Task simple-matching									**.067**	**.013**				
Task-conf. accuracy/bias (reg.-res.)											**.229**	**<.001**		
Self-concept accuracy/bias (reg.-res.)													**.344**	**<.001**
Explained variance	22.2%		10.9%		15.7%		18.6%		12.0%		16.8%		18.4%	
Unexplained variance (residual)	77.1%		86.0%		84.3%		81.3%		86.0%		82.9%		78.5%	
Unexplained variance (school)	.7%		3.2%		.0%		.2%		2.0%		.2%		3.1%	

Notes: Missing responses were estimated by expectation-maximisation. Standardised coefficients ('Std. Effect') and significance (p-values; 'Sig. (p)') are reported. Significant coefficients (p < .05) are highlighted in bold.

Section 7.3: Discussion

Students' science intentions were predicted by their confidence, attitudes, and other beliefs, in both surveys. Students' confidence expressed as self-concept beliefs was a strong predictor of their science intentions in the PISA 2006 survey, while confidence expressed as self-efficacy beliefs was a strong predictor in the 2014/2015 survey. The following sections discuss and contextualise these results, focusing on the students' various attitudes and wider factors (**Section 7.3.1**) and on their expressions of confidence and confidence accuracy/bias (**Section 7.3.2**), and highlight initial conclusions for the first research question (**Section 7.3.3**).

Section 7.3.1: Students' attitudes and further factors

In PISA 2006 for England, Year 11 students' science intentions were most strongly predicted (with standardised coefficients over .1) by the students' perceived utility of science, interest/enjoyment of science, personal value of science, and their science self-concept beliefs. In data collected in 2014/2015 in England, considered across students in Years 9, 10, and 11 while accounting for their varying ages, students' science intentions were most strongly predicted by the students' perceived utility of science, personal value of science, self-efficacy (confidence in their expected grades at GCSE/A-Level), subjective-norms/influences from parents, and the students' interest/enjoyment of science (although interest was just below the .1 magnitude threshold).

These results broadly confirmed the earlier hypotheses (**Section 4.2**) and accordingly confirmed findings seen within earlier studies, including the relevance of perceived utility of science, interest in science, and the influence of parents (DeWitt & Archer, 2015; Kjærnsli & Lie, 2011; Mujtaba & Reiss, 2014; Sjaastad, 2012; Wang & Degol, 2013). Concurrently, the results extended earlier studies through highlighting the strong predictive association between students' personal value of science to their identities and their science intentions. Students' science identities have generally been explored through qualitative perspectives, so that any relative impact compared to other factors could not easily be determined (Archer, et al., 2010; Aschbacher, Li, & Roth, 2010; Carlone & Johnson, 2007). Simple scales of agreement within questionnaires may not necessarily reflect the complex idea of identity, however, and the area can be considered in many ways. The students' personal value of science in PISA 2006 covered the personal relevance of science in understanding wider areas, relating to other people, and in being applied in various ways in the future (OECD, 2009a). The 2014/2015 survey directly covered theorised aspects of identity such as self-recognition ('Thinking scientifically is an important part of who I am') and conveying this to other people ('Being able to do science helps me show other people who I am') (Carlone & Johnson, 2007; Trautwein, et al., 2012; Wigfield & Eccles, 2000). Further research may need to explore how additional aspects of science identities could be measured or explored. Similarly, considering potential links between identity, interest, perceived difficulty or ability, and other areas of relevance to science education may be informative (Archer, et al., 2010; Archer, DeWitt, & Osborne, 2015); identity may also perhaps entail patterns of beliefs and associations between them, together with self-recognition and movements towards or away from particular ideals or roles.

The PISA 2006 survey and the 2014/2015 survey covered overlapping arrays of factors, but sometimes considered the same idea or factor in slightly different ways, as with the students' personal value of science. Similarities were nevertheless observed across the two sets of data, such as the importance of students' perceived utility of science and personal value of science. The slightly different measurement items plausibly reflected different perspectives onto similar underlying ideas, supported by the convergence of the results (Messick, 1995). However, the results notably differed across the two surveys regarding the relevance of students' interest, despite the similarities in measurement (which covered interest and enjoyment, such as via agreement or disagreement with 'I enjoy acquiring new knowledge in science' and 'I am interested in learning about science' in the PISA 2006 survey, and 'I enjoy learning science' and 'I am interested in the things I learn in science' in the 2014/2015 survey). Given that prior research has highlighted the strong relevance of students' interest to their choices (Holmegaard, Ulriksen, & Madsen, 2014; Jensen & Henriksen, 2015; Maltese, Melki, & Wiebke, 2014), some results may reflect the samples and may not necessarily generalise to wider students (and would need to be explored further); educational and social contexts also change over time, so some variability is also unavoidable.

The students' perceived utility value of science was the strongest predictor of science intentions for students in England in PISA 2006 and in the 2014/2015 survey, as also seen in earlier research studies (Mujtaba & Reiss, 2014). Within the expectancy-value model, utility value aims to consider the idea of science being valued due to indirect benefits (Eccles, 2009; Wigfield & Eccles, 2000); the similar idea of extrinsic (instrumental) motivation considers activities being undertaken to gain a separate or wider outcome (Deci & Ryan, 1985; Ryan & Deci, 2000). In the two surveys, utility value was measured with items such as 'Making an effort in science is worth it because it will help me in the work that I want to do later on' (OECD, 2009a). The 2014/2015 survey was supplemented by further items such as 'I need to do well in science to get the job I want' to be similarly comparable with further studies (Martin & Mullis, 2013).

For some students, studying science may be indirectly valued through helping them to widen their educational or career prospects: someone may aspire to a job that uses scientific but transferable skills in some way, rather than a job clearly within science itself. For other students who aspire towards a science career, it perhaps becomes difficult to distinguish utility as indirect value from utility reflecting some form of direct value; practically, it may be harder to disagree with items such as 'I need to do well in science to get the job I want' if the job that someone wants is within science. Future research may need to consider and separate students' perceptions of the indirect utility value of studying science (e.g. in providing transferable skills) from the direct utility value of studying science (e.g. in providing the skills and qualifications that are necessary to become a scientist). For example, considered from another perspective, it is even possible that someone may resolve to study science (perhaps due to interest or other factors), which then entails that they recognise and agree with the direct utility value of studying science (i.e. that they then need strong science qualifications to progress further); science intentions and direct utility value may be clearly associated, but causality may be reversed so that results might have less meaning.

Further benefit may also be gained from exploring why science qualifications and skills may be indirectly valued (or not valued). For example, science is often perceived as a relatively difficult subject (Archer, et al., 2010; Aschbacher, Li, & Roth, 2010; Bates, Pollard, Usher, & Oakley, 2009), and has indeed been found to require more generalised ability than other subjects to gain particular grades (Coe, 2008; He & Stockford, 2015). Prestige may then be gained from success in difficult examinations, for example, reflecting the idea that qualifications may embody various forms of indirect or exchange value (Archer, Dawson, DeWitt, Seakins, & Wong, 2015; Claussen & Osborne, 2013; Williams & Choudry, 2016). Additionally, it is possible that science and related skills are indirectly valued through facilitating higher potential income. Those with science-related degrees who work within science-related fields earn some of the highest salaries compared to other fields (Engineering UK, 2015; Royal Society, 2006; Walker & Zhu, 2013). Similarly, higher skills in science and mathematics have associated with increases in national productivity and economic returns across different countries (Hanushek & Kimko, 2000; Hanushek & Woessmann, 2012; OECD, 2015b), although students may not necessarily be aware of such wider contexts. Further research may then also need to explore why the indirect (or perhaps the direct) goal of studying science is valued or not and in what way (i.e. what, exactly, about a career in science or being a scientist is valued or attractive to students).

Considering wider factors, in both surveys, the students' gender (being a boy), the number of books at home, the level of fathers' education, and whether either parent worked within science were predictively associated with higher science intentions but only when the students' task-scores, confidence, and attitudes were not considered. While PISA 2006 did not measure students' ethnicity, reporting a South-Asian/Indian background (rather than a White background) predictively associated with higher intentions in the 2014/2015 data, even when modelling the students' various attitudes and other beliefs. These results matched earlier studies that have highlighted lower science intentions and/or choices for girls (Blickenstaff, 2005; Murphy & Whitelegg, 2006) and higher intentions/choices for students from ethnicities/backgrounds such as Indian (Archer & Francis, 2006; Connor, Tyers, Modood, & Hillage, 2004; Strand, 2007; Torgerson, et al., 2008). Nevertheless, considering further factors appears to be necessary in order to help explain such differences.

In general terms, these results supported the assumptions of the expectancy-value model of social-cognitive theory (Bandura, 1977, 1986, 1997; Eccles, 2009; Eccles, et al., 1983; Wigfield & Eccles, 2000). Essentially, any influences due to students' background and context have been theorised to be mainly mediated through students' attitudes and confidence expressed as self-concept or self-efficacy beliefs (Eccles, 2009). Further research may then need to consider, in more detail, how students' attitudes are formed or influenced. For example, parental education and fields of work may foster implicit (or perhaps explicit) dispositions within families that may influence students' interpretative frameworks and/or their attitudes and beliefs (Archer, Dawson, DeWitt, Seakins, & Wong, 2015; Bourdieu, 1984; Edgerton & Roberts, 2014). Families from various ethnicities/backgrounds, such as East-Asian/Chinese and Indian, have also often valued education and promoted particular careers for their children (Archer & Francis, 2006; Connor, Tyers, Modood, & Hillage, 2004; Strand, 2007; Torgerson, et al., 2008). University students' reported initial interest in science has also associated with parental encouragement and parents working within science (Dabney, Chakraverty, & Robert, 2013). Quantitative research may then need to encompass wider indicators, such as views from parents (and/or students' beliefs about their parents' preferences), in order to consider such areas.

The results from both surveys also highlighted that further factors could be directly predictive of students' intentions, but generally with lower magnitudes that perhaps entailed less practical relevance. For example, psychological studies have often considered factors such as motivational orientations to master learning, which have generally associated with higher interest and attainment (Elliot, 1999; Huang, 2011a; Jiang, Song, Lee, & Bong, 2014; Paulick, Watermann, & Nückles, 2013). The 2014/2015 survey highlighted that mastery orientations associated with higher science intentions, but only when the expectancy-value factors (i.e. including interest and attainment) were not modelled; when the expectancy-value factors were modelled, mastering learning became negatively predictive of science intentions, but was ultimately not predictive when the full array of factors was considered. The coefficient and significance changes suggested that complex patterns of mediation may occur (i.e. mastery orientations may reflect higher interest, attainment, or other factors, and/or may better predict these factors rather than science intentions); further research may then need to consider structural models of association in addition to predictive modelling of one outcome. Orientations to out-perform other students, however, remained negatively predictive of science intentions in the final predictive models, but with a minimal magnitude.

The predictive models covering PISA 2006 highlighted that different teaching approaches (i.e. interaction/discussion/debate, practical experiments, student-led investigations, and explaining/focusing on applied areas) had minimal (and sometimes negative) predictive associations with the students' science intentions, when also considering the students' various attitudes and beliefs. Teaching approaches have often been considered in the context of attainment, essentially in order to determine optimal or efficient practices (Cavagnetto, 2010; Schroeder, Scott, Tolson, Huang, & Lee, 2007). Teaching approaches and classroom experiences may influence students' interest and other attitudes to some extent, although it perhaps remains unclear how these would compare against other potential influences (Abrahams, 2009; Hampden-Thompson & Bennett, 2013; Wang, 2012). In general terms, practical work remains valued and recommended within science education, for example due to practical work being assumed to reflect the empirical nature of science itself (Royal Society, 2014). However, applying practical work in science education can be rationalised in various ways, and scientists, policy makers, teachers, and students may all have different interpretations of what practical work aims to achieve (Abrahams & Reiss, 2012; Abrahams, Reiss, & Sharpe, 2014; Hodson, 1993; Millar, 1998). Science education as being balanced between training future scientists (e.g. perhaps via practical laboratory work) while enhancing science literacy for everyone, and considering how to best undertake science teaching and learning, has been discussed for an extensive amount of time, but with no apparent resolution (Claussen & Osborne, 2013; Kimball, 1913;

Osborne & Dillon, 2008). In general, research and commentary on teaching and learning approaches can potentially lead to relatively clear practical impact (e.g. whether to apply a specific and tangible teaching approach or not), although may constrain enquiry and limit consideration of what students might prefer (e.g. teaching practices may be instead recommended following from contextualising learning theories, from assumptions about what scientists do or need to know, etc.). Further research may need to continue to consider how teaching associates with students' attitudes, and consider which areas of science and/or teaching/learning students favour; for example, some students might prefer learning about 'big ideas' or theories and/or applied areas of science and disfavour practical work or debate.

Students' studying strategies have also been variously considered in prior research studies, although to a lesser degree with secondary school students. For university students, the studying strategies of organisation, memorisation, and elaboration have generally been associated with higher attainment, with elaboration having the highest relative association (Credé & Phillips, 2011; Richardson, Abraham, & Bond, 2012). For secondary school students, however, memorisation strategies have been associated with lower attainment in mathematics (Chiu, Chow, & Mcbride-Chang, 2007; Echazarra, Salinas, Méndez, Denis, & Rech, 2016; Köller, 2001). The 2014/2015 survey highlighted that, in simple predictive models, self-regulated and elaboration studying strategies had the highest predictive associations with science intentions, higher than memorisation and control strategies. However, when the expectancy-value factors were also modelled (i.e. including students' attainment), only memorisation strategies remained predictive of higher science intentions, but with a very small magnitude. It perhaps remains unclear whether any particular studying strategies should be promoted, and if so, for what reasons. Researchers may also need to remain mindful of contextualisation; for example, memorisation may be perhaps unavoidable at secondary school and/or a pre-requisite to other learning strategies.

Many of the other areas considered in the 2014/2015 survey also appeared to be less relevant to students' science intentions. For example, the students' perceived control predictively associated with their science intentions, but at a minimal magnitude. Higher beliefs of perceived control (e.g. effort can lead to success in science, success depends on the individual, etc.) predicted slightly higher intentions when the expectancy-value factors were not modelled, but predicted slightly lower intentions when considering the comprehensive array of factors. It is possible that perceived control may influence students' attitudes more than intentions. Nevertheless, it perhaps remains important to continue to explore further potential predictors of students' science intentions, otherwise theoretical or analytical perspectives may become more prescriptive than informative. Analysis of Next Steps (formerly the Longitudinal Study of Young People in England), for example, has indeed highlighted that students' perceived control (believing in lack of success being due to the student i.e. an internal locus of control) has associated with studying science at A-Level, together with prior attainment and liking science (Department for Education, 2011). Given that some results from the 2014/2015 appear particular to the sample and may or may not generalise to other students, further research may be needed to help clarify the area.

Section 7.3.2: Students' confidence and confidence accuracy/bias

Focusing on the various expressions of confidence, students' self-concept beliefs, and to a lesser extent their self-efficacy beliefs, were predictive of their science intentions in the PISA 2006 survey. For the 2014/2015 survey, however, students' self-efficacy beliefs were predictive, while their self-concept beliefs were not. While both studies highlighted the relevance of self-efficacy and/or self-concept, the students' perceived utility of science and personal value of science ultimately had higher predictive associations with students' intentions.

These results broadly cohered with earlier studies. For example, students' self-concept beliefs have associated with science intentions in prior studies in England, which have also highlighted the importance of students' perceived utility of science (DeWitt & Archer, 2015; Mujtaba & Reiss, 2014). Other studies have also variously highlighted the associations between science intentions and self-concept or self-efficacy beliefs in various other countries (Bong, 2001b; Byars-Winston, Estrada, Howard, Davis, & Zalapa, 2010; Parker, et al., 2012; Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014).

The contrasting results from the two surveys may reflect and highlight the importance of contextual measurement. In the PISA 2006 survey, self-efficacy was measured as someone's expected capacity to undertake various applied science tasks that might be encountered in everyday life (e.g. 'Recognise the science question that underlies a newspaper report on a health issue', 'Interpret the scientific information provided on the labelling of food items') (OECD, 2009a). In the 2014/2015 survey, self-efficacy was measured as expected capacity to gain particular grades at GCSE and A-Level, and this had a relatively high predictive association with science intentions. In general terms, students' GCSE attainment has often been associated with selecting A-Level subjects (Department for Education, 2011, 2012; Homer, Ryder, & Banner, 2014); schools may require particular levels of GCSE attainment in order to study A-Level subjects (Department for Children, Schools and Families, 2009); and, additionally, for mathematics, students' intentions have associated with their predicted grades (Brown, Brown, & Bibby, 2008). The importance of self-efficacy, expressed as expected grades, plausibly cohered with and reflected such areas. Additionally, the higher association seen in the 2014/2015 survey may reflect that the measures of self-efficacy and intentions were both contextualised and measured at the subject level (Bandura, 1997; Bong, 1997; Pajares & Miller, 1995). Self-efficacy measured at the task level may be more relevant for task level outcomes (Bong, 1997; Jansen, Scherer, & Schroeders, 2015).

Alternately or additionally, the contrasting results may also follow from the different samples: the 2014/2015 sample was comparatively small and not necessarily nationally-representative; PISA 2006 covered a nationally-representative sample, although considered the situation as of that particular year. Regardless, given the partially unexpected results from the 2014/2015 survey (e.g. where self-concept and interest appeared less relevant), further research would realistically be necessary in order to help consider the relative influences of students' self-concept, self-efficacy, and any other expressions of confidence.

For both the surveys, indicators of accuracy/bias, formed via self-concept/task-score regression-residuals in the PISA 2006 survey and via task-level task-confidence/task-score difference-scores in the 2014/2015 survey, were predictive of students' science intentions; however, significance was lost in the 2014/2015 survey once students' wider attitudes and beliefs were also modelled as predictors. In PISA 2006, the accuracy/bias indicator remained predictive when it replaced the self-concept indicator, regardless of which other predictors were considered, which likely reflected the greater relative importance of self-concept beliefs in the PISA data. Additional models for PISA 2006 highlighted that, controlling for the various other predictors: task-scores and self-concept beliefs both positively predicted science intentions; task-scores and the accuracy/bias indicator both positively predicted science intentions (i.e. higher over-confidence, controlling for task-scores, predicted higher science intentions); self-concept positively predicted and the accuracy/bias indicator negatively predicted science intentions (i.e. higher over-confidence, controlling for self-concept beliefs, predicted lower science intentions). The same pattern was also seen in the 2014/2015 survey, but only when the task-level indicators were modelled alone with the students' background characteristics (otherwise significance was lost).

Accordingly, the two surveys provide reassurance that similar predictive patterns emerged, regardless of the methods of considering accuracy/bias, whether via differences in self-concept beliefs relative to task-scores across a sample, or via differences in self-reflective task-confidence explicitly linked to task-scores. However, the 2014/2015 survey ultimately highlighted that task-level self-reflective accuracy/bias was less directly relevant to subject-level studying intentions. This may unavoidably follow from task level measures being less generalizable or relevant to the subject level.

The additional indicators of students' task-level accuracy in the 2014/2015 survey were broadly similar to the task-level accuracy/bias indicator in that they were only predictive of science intentions when considered with the students' background characteristics, and were not predictive once the students' various other attitudes and beliefs were modelled. When considered with the students' background characteristics, higher sensitivity (being accurately confident when they had the right answer) predicted higher intentions, higher specificity (being accurately not confident when they had the wrong answer) predicted lower intentions, and higher simple-matching (both accurately knowing when answers were right and wrong) predicted slightly higher intentions. Despite earlier studies highlighting the importance of sensitivity and specificity (Schraw, Kuch, & Gutierrez, 2013; Schraw, Kuch, Gutierrez, & Richmond, 2014), it perhaps remains unclear how to best apply these ideas within science education. Considering task-level confidence accuracy may be directly meaningful for examinations and other assessment, and may require further exploration; for example, being able to accurately determine whether a task has been answered correctly or not may help students in deciding whether to consider the task again, which may help determine their allocation of effort and time.

Section 7.3.3: Conclusions

Research question one: which attitudes and motivational beliefs (including expressions of confidence) were the most relevant influences on students' science intentions?

Fundamentally, students' perceived utility of science and personal value of science were most strongly predictive of science intention across both PISA 2006 and the 2014/2015 survey. Students' confidence expressed as self-concept beliefs was a strong predictor of their science intentions in the PISA 2006 survey, while confidence expressed as self-efficacy beliefs (contextualised as confidence in their expected grades at GCSE/A-Level) was a strong predictor in the 2014/2015 survey. Indicators of accuracy/bias appeared to be directly predictive of science intentions if the underlying expression of confidence was also predictive.

However, indicators of confidence, attainment, and accuracy/bias could not be modelled together hence any wider implications were unavoidably less clear. The impact of accuracy/bias was then considered via grouping or clustering students, and whether students might consider their choices in different ways.

Section 8: Research question two

The second research question considered whether students with different degrees of confidence accuracy/bias exhibited different science intentions, attitudes, and beliefs. The results from the PISA 2006 survey (**Section 8.1**) and the 2014/2015 survey (**Section 8.2**) are described in turn, and then discussed and contextualised in **Section 8.3**.

Section 8.1: Results: PISA 2006

Predictive modelling could not easily reveal the impact of accuracy/bias on students' intentions (**Section 7**), so students were categorised according to their confidence accuracy/bias so that their expressed intentions, attitudes, and other beliefs could be considered and compared across the groups. Students were classified through different approaches in order to consider any similarities or differences: via conceptual accuracy/bias groups (**Section 8.1.1**) and via accuracy/bias clusters from latent-profile analysis (**Section 8.1.2**). Additionally, latent-profile analysis was used to form clusters based on students' intentions and key predictors (identified in **Section 7**), and differences in the cluster profiles were considered, including their degrees of accuracy/bias (**Section 8.1.3**).

Section 8.1.1: Accuracy/bias groups

Conceptual groups (above-average/below-average intersections)

Four conceptual accuracy/bias groups were formed to consider the intersection of below-average and above-average confidence and attainment (**Section 6.3**). These grouped (**Table 11**): those with above-average self-concept beliefs and above-average task-scores (33.2% of the considered students, i.e. broadly accurate with high confidence); those with above-average self-concept but below-average task-scores (18.2%, i.e. broadly over-confident); those with below-average self-concept but above-average task-scores (20.1%, i.e. broadly under-confident); and those with below-average self-concept and below-average task-scores (28.5%, i.e. broadly accurate with low confidence).

While such groups considered accuracy/bias broadly, they represented the conceptual ideas of under-confidence and over-confidence. The degree of exhibited (regression-residual) accuracy/bias also reflected the conceptual formation of the groups (means per group are shown in **Table 11**), although on average, students with accurately-high beliefs exhibited some degree of relative over-confidence (via the regression-residual indicator), while students with accurately-low beliefs exhibited some degree of relative under-confidence.

The students' science intentions differed when comparing all possible pairs of groups (means per group are shown in **Table 11** and differences are shown in **Table 12**). Since the analysis used the OECD factor-scores, which have no inherent meaning assigned to their values, standardised *z*-scores are reported (i.e. negative *z*-scores reflect below-average values while positive *z*-scores reflect above-average values, where values are given in standard deviations). Those with accurately-high beliefs reported the highest (above-average) science intentions, then over-confident students, then under-confident students, and then those with accurately-low beliefs. The magnitude of the overall difference in intentions across the groups was moderate: 19.8% of the variance in intentions could be attributed to the differences across the groups (η^2 (eta-squared) represents R^2 in this situation as a measure of effect size) (Cohen, 1988).

Group differences for the students' other attitudes and beliefs broadly followed the same pattern as seen in their intentions, including for the key predictors of students' intentions (the students' perceived utility of science, interest/enjoyment of science, personal value of science, and their science self-concept beliefs, as revealed in **Section 7**). The students' background factors showed different patterns across the groups, however. For example, there was no gender difference across the accurately-high and over-confident groups (both with above-average proportions of boys) or across the under-confident and accurately-low groups (both with below-average proportions of boys). For the other background indicators (number of books at home, either parent working in science, and parental educational level), fewer or no differences occurred between accurately-high and under-confident students and between over-confident and accurately-low students.

The students' self-concept beliefs and task-scores differed across the groups (by design given the method of forming the groups), but also empirically differed across all possible pairs of groups. For example, while the accurately-high and under-confident groups were both formed from students with 'above-average task-scores', the students in the accurately-high group nevertheless exhibited a higher magnitude of task-scores. Considering the patterns of magnitudes and differences across the groups, the pattern of science intentions and attitudes appeared to broadly reflect the students' self-concept beliefs (e.g. attitudes broadly descended across the accurately-high, then over-confident, then under-confident, then accurately-low groups), while the pattern of background indicators appeared to reflect the students' task-scores (e.g. accurately-high and under-confident students exhibited above-average and similar numbers of books at home).

On a wider level, it would be easy to then infer that holding accurately-high beliefs was the most beneficial case, in terms of the students' expressed attitudes and beliefs (which perhaps risks stating an intuitively obvious situation). The results nevertheless highlighted that over-confident students generally reported lower attitudes than accurately-high students, despite over-confident students holding potentially motivational above-average self-concept beliefs. Nevertheless, the artificial nature of these conceptual (above-average/below-average) accuracy/bias groups ensures that any findings cannot be definitive.

Perhaps problematically for science education, the above-average intentions held by the students with accurately-high beliefs (**Table 11**) may not necessarily entail strong aspirations towards science. As noted in the methods (**Section 5.2**), the OECD factor-scores were used for increased comparability against prior studies and national/international reports. While statistically sophisticated (OECD, 2009a), the factor-scores unavoidably draw attention away from the students' actual agreement or disagreement to the various underlying items. Calculated only for illustration (without estimating any missing responses), students across England in PISA 2006 on average 'disagreed' across the various items within the science intentions scale (mean (M) = 1.98; standard deviation (SD) = .80), each measured on a four-category scale with (1) ‘strongly disagree’, (2) ‘disagree’, (3) ‘agree’, and (4) ‘strongly agree’. The group of accurately-high students responded, on average, close to the mid-point of the agreement scale (M = 2.42; SD = .82), while the other groups responded lower within the range of disagreement and strong disagreement (over-confident M = 2.10, SD = .73; under-confident M = 1.69, SD = .65; accurately-low M = 1.60, SD = .60).

Table 11: PISA (England) 2006: conceptual groups (above-average/below-average self-concept/task-score intersections)

	Under-confident (U)		Accurately-low (L)		Accurately-high (H)		Over-confident (O)	
Item/factor (*z*-scores)	M	SD	M	SD	M	SD	M	SD
Self-concept accuracy/bias	-.92	.61	-.59	.74	.55	.70	.92	.67
Intentions	-.37	.86	-.49	.81	.55	.99	.17	.92
Gender (1=boy)	-.21	.97	-.19	.98	.19	.98	.13	.99
Books at home	.32	.88	-.37	.98	.39	.88	-.38	.97
Parent(s) in science (1=yes)	.09	1.07	-.17	.84	.18	1.12	-.17	.84
Mothers' education	.11	.95	-.19	1.04	.19	.95	-.13	1.04
Fathers' education	.07	.97	-.23	1.00	.24	.98	-.12	1.00
Task-score (PV1)	.61	.46	-.84	.61	.90	.59	-.77	.59
Self-concept	-.68	.59	-.86	.69	.82	.70	.60	.59
Self-efficacy (areas)	-.06	.75	-.68	.86	.70	.89	-.07	.84
Interest (various areas)	-.22	.84	-.48	1.07	.43	.76	.22	1.04
Interest value	-.32	.82	-.63	.86	.65	.87	.18	.88
Utility value	-.35	.87	-.47	.91	.49	.95	.22	.93
Personal value	-.30	.83	-.54	.88	.57	.96	.14	.87
General value	-.05	.83	-.51	.90	.50	1.00	-.03	.93
Science activities	-.21	.88	-.44	.91	.44	.94	.15	1.04
School career preparation	-.14	.91	-.45	.92	.46	.99	.07	.93
School career information	-.28	.92	-.26	1.00	.20	.97	.33	1.00
Teaching: interaction	-.23	.99	-.19	.97	.14	1.01	.25	.97
Teaching: activities	-.08	.83	-.27	1.07	.13	.90	.23	1.15
Teaching: investigations	-.34	.87	-.07	.97	-.02	.94	.42	1.13
Teaching: applications	-.21	.92	-.26	.97	.20	.99	.23	1.04
Group size (N, %)	949	20.1%	1348	28.5%	1573	33.2%	863	18.2%

Notes: Missing responses were estimated by expectation-maximisation. Means (M) and standard deviations (SD) are reported. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey.*

Table 12: PISA (England) 2006: conceptual groups (above-average/below-average self-concept/task-score intersections), group differences

	Overall group difference		Paired group differences, sig. (*p*)					
Item/factor	Sig. (*p*)	Size (η^2)	U-L	U-H	U-O	L-H	L-O	H-O
Self-concept accuracy/bias	**<.001**	.525	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Intentions	**<.001**	.198	**.008**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Gender (1=boy)	**<.001**	.034	1.000	**<.001**	**<.001**	**<.001**	**<.001**	.967
Books at home	**<.001**	.136	**<.001**	.444	**<.001**	**<.001**	1.000	**<.001**
Parent(s) in science (1=yes)	**<.001**	.026	**<.001**	.271	**<.001**	**<.001**	1.000	**<.001**
Mothers' education	**<.001**	.027	**<.001**	.367	**<.001**	**<.001**	1.000	**<.001**
Fathers' education	**<.001**	.036	**<.001**	**<.001**	**<.001**	**<.001**	.089	**<.001**
Task-score (PV1)	**<.001**	.665	**<.001**	**<.001**	**<.001**	**<.001**	.052	**<.001**
Self-concept	**<.001**	.579	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-efficacy (areas)	**<.001**	.291	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**
Interest (various areas)	**<.001**	.145	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Interest value	**<.001**	.272	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Utility value	**<.001**	.171	**.010**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Personal value	**<.001**	.207	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
General value	**<.001**	.156	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**
Science activities	**<.001**	.130	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
School career preparation	**<.001**	.130	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
School career information	**<.001**	.067	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**.008**
Teaching: interaction	**<.001**	.038	1.000	**<.001**	**<.001**	**<.001**	**<.001**	.083
Teaching: activities	**<.001**	.037	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	.113
Teaching: investigations	**<.001**	.057	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**
Teaching: applications	**<.001**	.050	1.000	**<.001**	**<.001**	**<.001**	**<.001**	1.000

Notes: Missing responses were estimated by expectation-maximisation. Significant p-values (p < .05) are highlighted in bold. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey.*

Conceptual groups (regression-residual accuracy/bias groups)

Students were also classified into conceptual (regression-residual) groups, through classifying the regression-residual indicator of accuracy/bias (and applying ±.5 standard deviations as group boundaries; **Section 6.3**); specifically, this process formed under-confident (28.9% of the considered students), accurately-evaluating (44.3%), and over-confident (26.8%) groups. Such groups have been considered in prior studies (Bouffard, Vezeau, Roy, & Lengelé, 2011; Gonida & Leondari, 2011; Narciss, Koerndle, & Dresel, 2011), and were directly considered in preliminary analysis. For greater insight, and to aid comparability across the various other groups and clusters of students, the accurately-evaluating students were then divided into those with above-average task-scores (23.6% of the overall number of considered students, i.e. those with accurately-high confidence) or below-average task-scores (20.6%, i.e. those with accurately-low confidence).

Cross-tabulating the regression-residual accuracy/bias groups and the previous 'above-average/below-average intersection' accuracy/bias groups highlighted that: the regression-residual under-confident group was mainly formed from those from the 'under-confident' (51.8%) and 'accurately-low' (46.5%) above-average/below-average groups; the regression-residual accurately-low group was mainly formed from the 'accurately-low' (71.6%) then the 'over-confident' (28.4%) groups; the regression-residual accurately-high group was mainly formed from the 'accurately-high' (78.5%) then the 'under-confident' (21.5%) groups; and the regression-residual over-confident group was mainly formed from those from the 'over-confident' (46.1%) and 'accurately-high' (52.8%) groups. Different grouping approaches unavoidably entail different classifications but there was some general similarity across both approaches, especially for the accurately-low and accurately-high groups.

The means per group and differences across groups are summarised in **Table 13** and **Table 14**. Following the nature of regression-residuals (Cohen & Cohen, 1984), under-confident, accurately-evaluating, and over-confident regression-residual groups would exhibit, on average, similar task-scores but different magnitudes of self-concept beliefs (i.e. their degrees of bias). Accordingly, the under-confident and over-confident groups exhibited similar task-scores. The accurately-low and accurately-high groups exhibited different task-scores only through being sub-groups of the wider 'accurately-evaluating' regression-residual group (which would otherwise exhibit average task-scores, similar to the under-confident and over-confident groups).

Considering the students' background, there were only differences between the under-confident and over-confident groups for gender (the under-confident group had a below-average proportion of boys, while the over-confident group had an above-average proportion) and level of fathers' education (under-confident students reported slightly below-average levels, while over-confident students reported slightly above-average levels); there were no differences for the number of books at home, parents working within science, and the level of mothers' education. There were no differences in parental education across the accurately-high and over-confident groups.

For the students' self-concept beliefs, and their other attitudes including their science intentions and its key predictors, under-confident students generally exhibited moderately below-average magnitudes, accurately-low students exhibited slightly below-average magnitudes, accurately-high students exhibited slightly above-average magnitudes, and over-confident students exhibited moderately above-average magnitudes. Fundamentally (**Table 13** / **Table 14**), 19.9% of the variance in science intentions followed from differences across the groups, highlighting a moderate effect.

On a wider level, considering these conceptual regression-residual accuracy/bias groups (and contrary to the results from the conceptual above-average/below-average accuracy/bias groups), it would be possible to infer that holding over-confident beliefs would be beneficial (rather than holding accurate beliefs), given that these students generally expressed the highest attitudes towards science (even when sub-dividing the accurately-evaluating students into those with accurately-low and those with accurately-high self-concept beliefs). Once again, however, the artificial nature of these regression-residual accuracy/bias groups still ensured that any findings cannot be definitive. Additionally, the above-average science intentions held by the over-confident students did not necessarily entail positive aspirations. For illustration (without estimating missing responses), for these regression-residual groups, over-confident students responded close to the mid-point of the underlying agreement scale (M = 2.43; SD = .84), while the other groups responded within the range of disagreement and strong disagreement (accurately-high M = 2.16, SD = .75; accurately-low M = 1.82, SD = .61; under-confident M = 1.53, SD = .61).

Table 13: PISA (England) 2006: conceptual groups (regression-residual accuracy/bias groups, ±.5 SD boundaries)

	Under-confident (U)		Accurately-low (L)		Accurately-high (H)		Over-confident (O)	
Item/factor (*z*-scores)	M	SD	M	SD	M	SD	M	SD
Self-concept accuracy/bias	-1.14	.62	.02	.28	-.01	.28	1.19	.60
Intentions	-.59	.82	-.19	.81	.23	.93	.56	1.02
Gender (1=boy)	-.24	.97	-.08	1.00	.04	1.00	.24	.97
Books at home	.03	.98	-.37	.98	.36	.87	-.02	1.02
Parent(s) in science (1=yes)	.01	1.01	-.17	.84	.12	1.09	.01	1.00
Mothers' education	-.01	1.01	-.15	1.02	.14	.94	.03	1.04
Fathers' education	-.06	1.01	-.19	.98	.18	.95	.08	1.05
Task-score (PV1)	.05	.90	-.78	.57	.77	.55	-.01	1.12
Self-concept	-1.08	.70	-.26	.32	.25	.35	1.12	.70
Self-efficacy (areas)	-.41	.94	-.42	.78	.36	.81	.47	1.04
Interest (various areas)	-.52	1.02	-.13	.97	.24	.71	.44	.97
Interest value	-.65	.89	-.23	.76	.29	.80	.61	.99
Utility value	-.58	.91	-.16	.86	.18	.88	.56	.97
Personal value	-.55	.90	-.25	.81	.21	.87	.58	.99
General value	-.32	.93	-.34	.85	.24	.88	.41	1.08
Science activities	-.43	.87	-.16	.96	.17	.90	.43	1.04
School career preparation	-.38	.98	-.21	.85	.19	.92	.41	1.03
School career information	-.38	1.01	-.03	.92	-.03	.91	.43	.98
Teaching: interaction	-.34	1.02	.00	.87	.02	.94	.32	1.04
Teaching: activities	-.27	1.00	-.08	.97	.04	.84	.28	1.09
Teaching: investigations	-.34	.90	.09	.97	-.14	.90	.35	1.09
Teaching: applications	-.35	.96	-.07	.93	.04	.91	.36	1.07
Group size (N, %)	1368	28.9%	976	20.6%	1119	23.6%	1270	26.8%

Notes: Missing responses were estimated by expectation-maximisation. Means (M) and standard deviations (SD) are reported. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey.*

Table 14: PISA (England) 2006: conceptual groups (regression-residual accuracy/bias groups, ±.5 SD boundaries), group differences

	Overall group difference		Paired group differences, sig. (*p*)					
Item/factor	Sig. (*p*)	Size (η^2)	U-L	U-H	U-O	L-H	L-O	H-O
Self-concept accuracy/bias	**<.001**	.756	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**
Intentions	**<.001**	.199	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Gender (1=boy)	**<.001**	.033	**.001**	**<.001**	**<.001**	**.026**	**<.001**	**<.001**
Books at home	**<.001**	.061	**<.001**	**<.001**	.760	**<.001**	**<.001**	**<.001**
Parent(s) in science (1=yes)	**<.001**	.010	**<.001**	**.029**	1.000	**<.001**	**<.001**	**.025**
Mothers' education	**<.001**	.009	**.004**	**.002**	1.000	**<.001**	**<.001**	.079
Fathers' education	**<.001**	.018	**.010**	**<.001**	**.001**	**<.001**	**<.001**	.121
Task-score (PV1)	**<.001**	.275	**<.001**	**<.001**	.252	**<.001**	**<.001**	**<.001**
Self-concept	**<.001**	.686	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-efficacy (areas)	**<.001**	.176	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**.017**
Interest (various areas)	**<.001**	.146	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Interest value	**<.001**	.246	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Utility value	**<.001**	.190	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Personal value	**<.001**	.198	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
General value	**<.001**	.111	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Science activities	**<.001**	.114	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
School career preparation	**<.001**	.103	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
School career information	**<.001**	.090	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**
Teaching: interaction	**<.001**	.062	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**
Teaching: activities	**<.001**	.043	**<.001**	**<.001**	**<.001**	**.038**	**<.001**	**<.001**
Teaching: investigations	**<.001**	.072	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Teaching: applications	**<.001**	.071	**<.001**	**<.001**	**<.001**	**.042**	**<.001**	**<.001**

Notes: Missing responses were estimated by expectation-maximisation. Significant p-values ($p < .05$) are highlighted in bold. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey.*

Section 8.1.2: Accuracy/bias clusters

Different methods of grouping students entailed different ideas of 'under-confidence' and 'over-confidence'. For example, applying the intersection between below-average and above-average self-concept and task-scores assumed and entailed that under-confident students would exhibit above-average task-scores and below-average self-concept beliefs; applying regression-residuals as an indicator of accuracy/bias assumed and entailed that under-confident students would exhibit average task-scores and below-average self-concept beliefs. Both approaches were valid in covering the underlying ideas of under-confidence (lower confidence than attainment) and over-confidence (higher confidence than attainment) applied as an analytical perspective, but relied on particular conceptual definitions that varied per approach. Neither approach could comprehensively consider the overall prevalence of under-confidence or over-confidence across England. Fundamentally, while various conceptual accuracy/bias groups could be formed, it remained unclear whether such groups would be naturally observed or would naturally emerge from data.

For greater insight, latent-profile analysis (**Section 6.3**) was undertaken using the students' reported self-concept beliefs and task-scores (the first plausible-value). This identified clusters of students, each with distinct magnitudes of self-concept beliefs and task-scores; implicitly, each cluster then had a distinct magnitude of accuracy/bias. The clusters could then highlight any particular tendencies of accuracy/bias and their prevalence (i.e. the cluster sizes).

Essentially, the clusters were indirectly formed on the accuracy/bias of the students' self-concept beliefs, but without directly calculating an indicator of accuracy/bias and/or forming artificial groups. The latent-profile analysis therefore applied accuracy/bias as an analytical perspective while also testing the analytical perspective itself. One cluster (i.e. the whole sample considered together) might fit the data better than numerous clusters, which would highlight that considering accuracy/bias would be less meaningful.

Operationally, multiple-imputation was not supported by the latent-profile analysis software (Vermunt & Magidson, 2013), so only the first plausible-value was considered in the final models. Preliminary analysis highlighted that similar cluster profiles and sizes resulted from modelling self-concept with each of the five plausible-values (i.e. the different plausible-values were essentially interchangeable). Additionally, preliminary latent-profile analysis with all five plausible-values (and not including self-concept beliefs), and with all five plausible-values and self-concept beliefs, resulted in similar cluster profiles and proportions. Conceptually, modelling all five plausible-values together with students' self-concept beliefs did not form groups based on the accuracy/bias of students' beliefs: the differences between the five plausible-values themselves were given equal relevance to the differences between the plausible-values and the students' self-concept beliefs; empirically, there were more indicators of task-score than self-concept, and the clusters were then essentially formed by considering task-score. As a further methodological check, preliminary analysis produced similar results (i.e. cluster profiles and proportions) regardless of how the students' self-concept beliefs and task-scores were operationalised (e.g. using the OECD factors, using re-calculated simple-average factors, etc.). As before, the final analysis used the OECD factors, to enhance contextualisation against any existing OECD material and related research studies.

Identifying the number of clusters to consider

Modelling one to ten clusters highlighted relatively linear improvements in the model information criteria (AIC, BIC, and sample-size adjusted BIC) as the number of modelled clusters increased (**Table 15**); lower information criteria represented better fit to the data. There was, therefore, no clearly 'ideal' number of clusters to consider. Nevertheless, the improvements in the information criteria highlighted that considering multiple clusters rather than one single cluster (i.e. the entire sample) was indeed plausible (and empirically validated).

Different insights may be gained from considering different numbers of clusters. However, higher numbers of modelled clusters may entail numerous smaller clusters, which may be too small for reliable predictive modelling. Higher numbers of clusters also become increasingly difficult to compare and briefly summarise (i.e. the number of comparisons across pairs of clusters increases substantially). The various cluster sizes and profiles, and changes across the different numbers of clusters being modelled, were then considered.

The automatic labelling of clusters (i.e. A, B, C, etc.) broadly followed their descending size order (although with some exceptions for some of the smaller clusters). Some clusters of students naturally emerged regardless of the numbers of clusters being modelled. For example, cluster B from the three-cluster model broadly emerged as a distinct cluster in the subsequent models that identified four to ten clusters, for example being classified as cluster C in the four-cluster model and cluster B in the five-cluster model. Accordingly, considering the four-cluster model, cluster C then appeared as a distinct cluster in all the subsequent models (i.e. when identifying five to ten clusters), such as emerging as cluster B in the five-cluster model. Considering the five-cluster model, clusters B, C, and E then appeared as distinct clusters in all the subsequent models; cluster D from the five-cluster model also broadly emerged in models with eight, nine, and ten clusters (with the addition of a small number of other students). Considering subsequent changes from five clusters to six, from six to seven, and so on, generally involved changes to two clusters each time, although the changes were not necessarily hierarchical. Clusters could reform and did not always involve one cluster simply dividing into two further clusters. For example, cluster A from the five-cluster model mainly divided into two clusters (A and B) in the six-cluster model, but cluster A in the six-cluster model also included students from cluster D from the five-cluster model.

Accordingly, and considered broadly, four or five clusters appeared to be plausible foci for analysis, balancing interpretability, complexity, and the number of students per cluster (which facilitated reliable predictive modelling per cluster). More clusters would provide greater insight and fit to the data, but would be far less practical to interpret. One of the four clusters from the four-cluster model, and three of the five clusters from the five-cluster model, emerged again when identifying higher numbers of clusters, highlighting that these appeared to be key groups to consider. Four clusters matched the number of conceptual groups, while five clusters potentially provided greater insight and empirical fit to the data. Ultimately, for brevity and consistency across all areas of the analysis, and to facilitate predictive modelling via larger cluster sizes (**Section 9**), four clusters were focused on (five clusters were considered in preliminary analysis and provided similar insights; see **Appendix 9** for illustration).

Accuracy/bias clusters from latent-profile analysis (four clusters)

The clusters from the four-cluster model had varying sizes (**Table 16**), from cluster A (55.9% of the considered students), B (17.8%), C (15.9%), to D (10.4%). Tabulating the clusters against the conceptual above-average/below-average accuracy/bias groups in order to help contextualise the clusters highlighted that cluster A, the largest cluster, was mainly composed of 'accurately-low' students (31.1% of the cluster) then 'under-confident' students (30.9%). Cluster B was mainly composed of 'accurately-low' students (62.5%) then 'over-confident' students (20.6%). Cluster C was mainly composed of 'accurately-high' students (65.5%) then 'over-confident' students (34.5%), while cluster D was primarily composed of 'accurately-high' students (95.7%).

Alternately, tabulating the clusters against the conceptual regression-residual groups highlighted that cluster A was mainly composed of 'accurately-low' (33.5%), 'accurately-high' (30.7%), and 'under-confident' students (30.6%). Cluster B was mainly composed of 'under-confident' students (66.3%). Cluster C was mainly composed of 'over-confident' students (60.9%) then 'accurately-high' students (39.1%), while cluster D was almost exclusively composed of 'over-confident' students (97.8%).

Means per cluster are summarised in **Table 16** and differences across the clusters are summarised in **Table 17**. The students' self-concept beliefs, task-scores, and accuracy/bias differed across all pairs of clusters (as expected, given the process). Students in cluster B reported the lowest self-concept beliefs and task-scores, but exhibited under-confidence via the regression-residual indicator (reflecting relative comparisons against other students, given the student's own task-scores). Conversely, students in cluster D reported the highest self-concept beliefs and task-scores, but exhibited relative over-confidence. Accordingly, given the constraints of the PISA design, it was difficult to definitively characterise each cluster in terms of their accuracy/bias: the regression-residual indicator reflected relative accuracy/bias, not necessarily absolute or self-reflective accuracy/bias.

The students' reported attitudes and motivational beliefs (**Table 16**) generally differed across all pairs of clusters (**Table 17**), with small numbers of exceptions. The students' reported attitudes and beliefs appeared broadly proportional to their self-concept beliefs and task-scores when compared across the clusters. For example, those in cluster B (with the lowest self-concept and task-scores) consistently reported the lowest attitudes and those in cluster D (with the highest self-concept and task-scores) consistently reported the highest, including for their science intentions. Essentially, students in cluster B exhibited moderately below-average attitudes, students in cluster A exhibited slightly below-average attitudes, students in cluster C exhibited moderately above-average attitudes, and students in cluster D exhibited highly above-average attitudes. The students' background and other reported characteristics also varied across the clusters, for example with above-average proportions of boys being present in cluster D (with the highest attitudes and beliefs), who also reported the highest levels of books at home, parental education, and parents working within science.

Fundamentally, the students' reported science intentions, and key predictors, differed in reported magnitude across all pairs of clusters; 20.1% of the variance in intentions could be attributed to the differences across the clusters (**Table 17**). For further illustration (without estimating missing responses), considering science intentions contextualised against the original agreement-scale with (1) 'strongly disagree', (2) 'disagree', (3) 'agree', and (4) 'strongly agree', students in cluster B responded the lowest, on average between strong disagreement and disagreement (M = 1.56, SD = .75), while students in cluster A responded on average with disagreement (M = 1.87, SD = .66). Students in cluster C responded slightly below the mid-point of the scale (M = 2.31, SD = .73), while students in cluster D held positive intentions, closer to agreement (M = 2.82, SD = .84).

As a wider point, and somewhat curiously, cluster C was formed from students who exhibited essentially the same self-concept beliefs, with little to no variance in self-concept within the cluster (**Table 16**). Considering the frequencies of response-categories for the items within the self-concept factor, students in cluster C primarily ‘agreed’ with all the items (on the four-category scale with ‘strongly disagree’, ‘disagree’, ‘agree’, and ‘strongly agree). It was unclear whether this was a particular acquiescence or any other response tendency (e.g. indifference, lower self-reflection, or rushed answers). The self-concept items were the last questions on the PISA 2006 student questionnaire, but the students’ responses to the preceding questions showed more variation (i.e. and did not necessarily suggest simply ‘agreeing’ due to fatigue or having to respond swiftly due to running out of assessment time, unless this only became an issue for the final questionnaire page). A pattern of responses may not necessarily entail that the students were somehow answering ‘problematically’ or were somehow at fault; the four-category response scale may not have been sufficient to distinguish differences in the students’ confidence, for example, compared to six-category or other response scales. Nevertheless, the clusters may have reflected tendencies other than accuracy/bias.

As another wider point, the clusters were formed via latent-profile analysis, implicitly considering accuracy/bias via forming clusters on the students’ self-concept beliefs and task-scores. Further insight could potentially be gained through considering what items/factors predicted membership of the different clusters, other than self-concept beliefs and task-scores. Such predictors may help explain why particular tendencies towards accuracy/bias may have arisen, although the predictors would also help explain the particular cluster magnitudes of self-concept beliefs and task-scores. It may then be harder to practically interpret the results, and separate antecedents of accuracy/bias from antecedents of self-concept beliefs or task-scores. Fundamentally, given the focus on students’ intentions, such areas would only offer further contextualisation and are outside the scope of the current research questions; nevertheless, this may offer further potential for future research.

Table 15: PISA (England) 2006: accuracy/bias clusters, model information criteria

Clusters	BIC	AIC	SA-BIC
1	69884.03	69858.18	69871.32
2	69363.75	69305.59	69335.15
3	67090.22	66999.75	67045.73
4	66213.12	66090.34	66152.75
5	64468.65	64313.55	64392.38
6	63833.33	63645.93	63741.18
7	61948.03	61728.31	61839.99
8	58768.84	58516.81	58644.91
9	57615.08	57330.73	57475.26
10	55006.01	54689.36	54850.31

Notes: The Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and sample-size adjusted Bayesian Information Criterion (SA-BIC) from the respective solutions for the numbers of clusters are reported. Lower information criteria reflected better fit to the data.

Table 16: PISA (England) 2006: accuracy/bias clusters (four clusters)

	Cluster A		Cluster B		Cluster C		Cluster D	
Item/factor (*z*-scores)	M	SD	M	SD	M	SD	M	SD
Self-concept accuracy/bias	-.24	.49	-.65	1.53	.65	.36	1.45	.59
Intentions	-.13	.86	-.56	.97	.43	.89	1.02	1.00
Gender (1=boy)	-.08	1.00	-.17	.98	.14	.99	.41	.92
Books at home	.01	.98	-.34	1.00	.13	1.00	.44	.90
Parent(s) in science (1=yes)	-.01	.99	-.16	.86	.04	1.03	.26	1.15
Mothers' education	-.01	.98	-.13	1.07	.04	1.00	.27	.99
Fathers' education	-.04	.98	-.15	1.06	.10	.98	.35	1.05
Task-score (PV1)	.04	.80	-.88	.89	.30	1.01	1.08	.66
Self-concept	-.23	.40	-.93	1.36	.71	.00	1.75	.53
Self-efficacy (areas)	-.12	.77	-.70	.98	.48	.95	1.17	.92
Interest (various areas)	-.04	.83	-.60	1.28	.34	.87	.76	.82
Interest value	-.11	.79	-.72	1.09	.43	.80	1.20	.88
Utility value	-.12	.86	-.53	1.09	.38	.90	.93	.95
Personal value	-.12	.83	-.59	1.06	.37	.84	1.11	.96
General value	-.09	.87	-.50	1.02	.26	.94	.98	1.02
Science activities	-.10	.91	-.42	1.03	.31	.94	.83	.94
School career preparation	-.08	.88	-.48	1.08	.27	.94	.89	.96
School career information	-.08	.92	-.27	1.14	.26	.94	.52	1.04
Teaching: interaction	-.04	.92	-.23	1.16	.13	.95	.37	1.13
Teaching: activities	-.04	.91	-.21	1.25	.13	.98	.29	1.00
Teaching: investigations	-.11	.94	.02	1.13	.22	1.01	.09	1.01
Teaching: applications	-.07	.91	-.26	1.15	.17	1.00	.48	1.09
Cluster size (N, %)	2644	55.9%	841	17.8%	754	15.9%	494	10.4%

Notes: Missing responses were estimated by expectation-maximisation. Means (M) and standard deviations (SD) are reported. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey.*

Table 17: PISA (England) 2006: accuracy/bias clusters (four clusters), cluster differences

	Overall group difference		Paired group differences, sig. (p)					
Item/factor	Sig. (p)	Size (η^2)	A-B	A-C	A-D	B-C	B-D	C-D
Self-concept accuracy/bias	**<.001**	.391	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Intentions	**<.001**	.201	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Gender (1=boy)	**<.001**	.029	.194	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Books at home	**<.001**	.043	**<.001**	**.026**	**<.001**	**<.001**	**<.001**	**<.001**
Parent(s) in science (1=yes)	**<.001**	.012	**.002**	1.000	**<.001**	**.001**	**<.001**	**.002**
Mothers' education	**<.001**	.011	**.017**	1.000	**<.001**	**.004**	**<.001**	**<.001**
Fathers' education	**<.001**	.019	**.049**	**.004**	**<.001**	**<.001**	**<.001**	**<.001**
Task-score (PV1)	**<.001**	.280	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-concept	**<.001**	.563	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-efficacy (areas)	**<.001**	.273	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Interest (various areas)	**<.001**	.142	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Interest value	**<.001**	.275	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Utility value	**<.001**	.167	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Personal value	**<.001**	.218	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
General value	**<.001**	.158	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Science activities	**<.001**	.123	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
School career preparation	**<.001**	.135	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
School career information	**<.001**	.055	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Teaching: interaction	**<.001**	.027	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Teaching: activities	**<.001**	.020	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**.028**
Teaching: investigations	**<.001**	.015	**.006**	**<.001**	**<.001**	**<.001**	1.000	.128
Teaching: applications	**<.001**	.042	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**

*Notes: Missing responses were estimated by expectation-maximisation. Significant p-values ($p < .05$) are highlighted in bold. Science intentions and their key predictors (identified in **Section 7**) are shaded in grey.*

Section 8.1.3: Intention/attitude clusters

For students in England in PISA 2006, science intentions were most strongly predicted by their perceived utility of science, interest/enjoyment of science, personal value of science, and their science self-concept beliefs (**Section 7**). These findings informed further analysis to provide a complementary perspective onto the prevalence of confidence accuracy/biases: latent-profile analysis was undertaken using the students' science intentions and these key predictive factors (perceived utility of science, interest/enjoyment of science, personal value of science, and science self-concept beliefs) to reveal further (intention/attitude) clusters of students. The results could then consider whether the students' task-scores, and hence their self-concept accuracy/bias, differed across any emerging clusters. Essentially, this explored whether confidence accuracy/bias actually differed across clusters of students who might be of particular contextual relevance to science education.

Additionally, the intention/attitude cluster profiles and proportions could help quantify the proportions of students likely to aspire towards science. Previous research has highlighted that students in England have, on average, generally considered science to be fairly interesting and relevant for careers, but have not necessarily aspired towards science (DeWitt, Archer, & Osborne, 2014; Jenkins & Nelson, 2005). However, the appearance of positive (or negative) attitudes may be misleading if the sample-average results from a small cluster of students with high attitudes while the majority of students hold lower attitudes, or from any other combinations of clusters.

Identifying the number of clusters to consider

Modelling one to ten clusters revealed improvements in the model information criteria as the number of modelled clusters increased, but highlighted diminishing improvements after five clusters (**Table 18**). Tabulations of the various clusters highlighted that some consistent clusters were increasingly revealed, regardless of the number of clusters modelled.

Four clusters were considered for contextualisation and comparability with the earlier results from the conceptual accuracy/bias groups and accuracy/bias clusters. However, considering five clusters (or more) would be beneficial in future research. Preliminary analysis essentially highlighted that the four-cluster and five-cluster solutions revealed similar profiles and proportions for the clusters exhibiting the lowest and highest attitudes, while the five-cluster solution provided greater differentiation of those with moderately positive/negative attitudes (**Appendix 9** shows the five-cluster solution). For the purposes of addressing the research question, accuracy/bias differed across both the four-cluster and the five-cluster intention/attitude cluster solutions.

Intention/attitude clusters from latent-profile analysis (four clusters)

The intention/attitude clusters from the four-cluster model had varying sizes (**Table 19**), from cluster A (41.5% of the considered students), B (24.4%), C (22.3%), to D (11.7%).

Tabulating the clusters against the conceptual above-average/below-average accuracy/bias groups highlighted that cluster A was broadly composed of all groups. Cluster B was mainly composed of 'accurately-high' students (59.9%) then 'over-confident' students (24.3%). Cluster C was mainly composed of 'accurately-low' students (48.8%) then 'under-confident' students (28.5%). Cluster D was mainly composed of 'accurately-high' students (65.0%) and 'over-confident' students (20.8%).

Tabulating the clusters against the conceptual regression-residual groups highlighted that cluster A was mainly composed of 'under-confident' students (34.2%), 'accurately-low' (28.6%), and 'accurately-high' (23.0%) students. Cluster B was mainly composed of 'over-confident' (46.9%) and 'accurately-high' students (34.0%). Cluster C was mainly composed of 'under-confident' (55.1%) then 'accurately-low' students (21.4%). Cluster D was mainly composed of 'over-confident' (60.3%) then 'accurately-high' students (24.4%).

The intention/attitude clusters did not clearly correspond to the earlier accuracy/bias clusters. The majority of the students from the accuracy/bias cluster A (54.0% of the cluster) were assigned to the intention/attitude cluster A. The accuracy/bias cluster B was mostly assigned across intention/attitude cluster C (48.7%) and cluster A (34.5%). The accuracy/bias cluster C was mostly assigned across intention/attitude cluster B (47.1%) and cluster A (28.3%). The accuracy/bias cluster D was mostly assigned across intention/attitude cluster B (48.0%) and cluster D (42.7%).

Considered broadly, the students' reported attitudes and motivational beliefs significantly differed across all potential pairs of clusters (**Table 19** / **Table 20**). This was expected, given that latent-profile analysis would likely ensure that factor averages differ across any emerging clusters for the modelled factors. Nevertheless, the students' other beliefs also similarly varied across the clusters, while their background characteristics also varied to some extent. Notably, the students' task-scores, and the magnitudes of regression-residual indicators of self-concept accuracy/bias, also differed across all pairs of groups.

Essentially, the students' reported science attitudes and beliefs were moderately to highly below-average in cluster C (22.3% of the modelled students), slightly below-average in cluster A (41.5%), moderately above-average in cluster B (24.4%), and highly above-average in cluster D (11.7%). Students' task-scores were observed with a similar pattern, increasing across clusters C, A, B, to D. Similarly, the students' self-concept accuracy/bias progressed from relative (regression-residual) under-confidence to relative over-confidence across clusters C, A, B, to D.

For further illustration (without estimating missing responses), considering science intentions contextualised against the original agreement-scale with (1) 'strongly disagree', (2) 'disagree', (3) 'agree', and (4) 'strongly agree', students in cluster C responded on average with strong (and rather clearly unvarying) disagreement (M = 1.00, SD = .00). Students in cluster A responded with disagreement (M = 1.86, SD = .37), while students in cluster B responded around the scale mid-point (M = 2.54, SD = .56). Students in cluster D responded positively with agreement (M = 3.11, SD = .68). However, as above, cluster D was the smallest with only 11.7% of the considered students.

These intention/attitude clusters (**Table 19** / **Table 20**) and the previous accuracy/bias clusters (**Table 16** / **Table 17**) varied in their proportions, and cross-tabulating the intention/attitude and the accuracy/bias clusters revealed less correspondence. However, a common tendency was nevertheless observed across both sets of clusters. A minority of students (i.e. 10.4% embodying accuracy/bias cluster D in **Table 16**, and 11.7% embodying intention/attitude cluster D in **Table 19**) emerged with highly above-average self-concept and task-scores (with relative ‘over-confidence’ from the regression-residual indicator of accuracy/bias), and with highly above-average attitudes and beliefs regarding science. Similarly, a greater proportion of students were observed with less-extreme but below-average self-concept, attitudes, and task-scores, and apparent relative ‘under-confidence’ (i.e. cluster A in **Table 16** and in **Table 19**).

Considered briefly for further insight (see **Appendix 9** for result tables), the five-cluster solution similarly highlighted that students’ task-scores, and the associated regression-residual indicators of accuracy/bias in self-concept beliefs, differed across all pairs of clusters; their magnitudes broadly corresponded to their reported attitudes and beliefs, as seen in the four-cluster solution. In the five-cluster solution, the students’ reported intentions, attitudes, and motivational beliefs were moderately to highly below-average in cluster B (22.4% of the modelled students), slightly to moderately below-average in cluster D (16.7%), slightly below-average in cluster C (21.1%), slightly to moderately above-average in cluster A (28.2%), and highly above-average in cluster E (11.5%). For further illustration (without estimating missing responses), considering science intentions on the original agreement-scale, students in cluster B responded with unvarying strong disagreement (M = 1.00, SD = .00). Students in cluster D responded with disagreement (M = 1.49, SD = .20), as did students in cluster C (M = 2.00, SD = .00). Students in cluster A responded around the mid-point (M = 2.57, SD = .50), while students in cluster E responded positively with agreement (M = 3.14, SD = .66). Essentially, the same inferences could be drawn from the four-cluster and the five-cluster solutions.

Table 18: PISA (England) 2006: intention/attitude clusters, model information criteria

Clusters	BIC	AIC	SA-BIC
1	63421.47	63356.97	63389.70
2	53764.20	53628.74	53697.47
3	47855.97	47649.56	47754.28
4	44339.85	44062.48	44203.21
5	39861.06	39512.74	39689.47
6	38411.80	37992.52	38205.26
7	37509.57	37019.34	37268.07
8	35511.55	34950.37	35235.10
9	34954.52	34322.38	34643.12
10	33711.89	33008.80	33365.53

Notes: The Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and sample-size adjusted Bayesian Information Criterion (SA-BIC) from the respective solutions for the numbers of clusters are reported. Lower information criteria reflected better fit to the data.

Table 19: PISA (England) 2006: intention/attitude clusters (four clusters)

	Cluster A		Cluster B		Cluster C		Cluster D	
Item/factor (*z*-scores)	M	SD	M	SD	M	SD	M	SD
Self-concept accuracy/bias	-.21	.75	.57	.79	-.67	1.02	.84	.94
Intentions	-.10	.48	.73	.65	-1.33	.00	1.36	.78
Gender (1=boy)	-.06	1.00	.19	.98	-.17	.98	.03	1.00
Books at home	-.05	.98	.19	.96	-.16	1.04	.18	.97
Parent(s) in science (1=yes)	-.08	.93	.10	1.07	-.07	.94	.18	1.12
Mothers' education	-.04	.99	.08	1.01	-.05	1.02	.09	1.04
Fathers' education	-.06	.98	.18	.98	-.16	1.01	.19	1.06
Task-score (PV1)	-.11	.92	.35	1.00	-.29	.89	.53	.97
Self-concept	-.25	.69	.66	.75	-.75	.99	.97	.94
Self-efficacy (areas)	-.19	.82	.45	.93	-.44	1.01	.70	.96
Interest (various areas)	-.10	.72	.52	.66	-.80	1.17	.83	.84
Interest value	-.25	.69	.71	.69	-.85	.87	1.07	.90
Utility value	-.31	.57	.51	.55	-.91	.84	1.77	.00
Personal value	-.31	.61	.69	.73	-.81	.85	1.23	.87
General value	-.27	.79	.48	.95	-.44	.94	.84	1.02
Science activities	-.21	.87	.56	.87	-.61	.83	.76	.94
School career preparation	-.19	.83	.38	.90	-.49	1.02	.87	1.00
School career information	-.10	.88	.32	.89	-.51	1.07	.63	1.00
Teaching: interaction	-.06	.93	.22	.98	-.31	1.05	.29	1.06
Teaching: activities	-.09	.93	.24	.98	-.25	1.09	.28	.98
Teaching: investigations	-.06	.97	.21	1.04	-.26	.94	.16	1.05
Teaching: applications	-.11	.89	.27	.98	-.35	1.05	.46	1.04
Cluster size (N, %)	1941	41.5%	1143	24.4%	1044	22.3%	549	11.7%

*Notes: Missing responses were estimated by expectation-maximisation. Means (M) and standard deviations (SD) are reported. Science intentions and their key predictors (identified in **Section 7**) are shaded in grey; the clusters were formed on these factors.*

Table 20: PISA (England) 2006: intention/attitude clusters (four clusters), cluster differences

Item/factor	Overall group difference		Paired group differences, sig. (*p*)					
	Sig. (*p*)	Size (η^2)	A-B	A-C	A-D	B-C	B-D	C-D
Self-concept accuracy/bias	**<.001**	.282	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Intentions	**<.001**	.733	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Gender (1=boy)	**<.001**	.017	**<.001**	**.022**	.367	**<.001**	**.010**	**.001**
Books at home	**<.001**	.019	**<.001**	**.028**	**<.001**	**<.001**	1.000	**<.001**
Parent(s) in science (1=yes)	**<.001**	.010	**<.001**	1.000	**<.001**	**.001**	.840	**<.001**
Mothers' education	**.001**	.003	**.010**	1.000	.079	**.018**	1.000	.082
Fathers' education	**<.001**	.019	**<.001**	.085	**<.001**	**<.001**	1.000	**<.001**
Task-score (PV1)	**<.001**	.088	**<.001**	**<.001**	**<.001**	**<.001**	**.001**	**<.001**
Self-concept	**<.001**	.358	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-efficacy (areas)	**<.001**	.166	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Interest (various areas)	**<.001**	.291	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Interest value	**<.001**	.435	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Utility value	**<.001**	.643	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Personal value	**<.001**	.473	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
General value	**<.001**	.210	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Science activities	**<.001**	.244	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
School career preparation	**<.001**	.188	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
School career information	**<.001**	.131	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Teaching: interaction	**<.001**	.043	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Teaching: activities	**<.001**	.041	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Teaching: investigations	**<.001**	.030	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Teaching: applications	**<.001**	.075	**<.001**	**<.001**	**<.001**	**<.001**	**.001**	**<.001**

Notes: Missing responses were estimated by expectation-maximisation. Significant p-values (p < .05) are highlighted in bold. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey; the clusters were formed on these factors.*

Section 8.2: Results: 2014/2015 survey

The students surveyed in 2014/2015 were classified via conceptual accuracy/bias groups (**Section 8.2.1**) and via accuracy/bias clusters from latent-profile analysis (**Section 8.2.2**). As with PISA 2006, additional latent-profile analysis was used to form clusters based on students' intentions and key predictors (identified in **Section 7**), and differences in the cluster profiles were then considered, including their degrees of accuracy/bias (**Section 8.2.3**).

Section 8.2.1: Accuracy/bias groups

Conceptual groups (difference-score groups with accurately-high and accurately-low groups)

Following traditional approaches (Gonida & Leondari, 2011), the difference-score indicator of task-level accuracy/bias was used to initially classify students as under-confident, accurate, or over-confident, relative to their academic year (i.e. the indicator was standardised as a *z*-score per year, and ±.5 standard deviations were used as the group boundaries, as described in **Section 5.3**). Cross-tabulating these difference-score accuracy/bias groups (under-confident, accurate, and over-confident, relative to the students' academic year) with indicators of performance (below-average and above-average task-scores, again relative to the students' academic year) highlighted that the majority of under-confident students exhibited above-average task-scores while the majority of over-confident students exhibited below-average task-scores. For additional insight, the accurate group was then divided into those with above-average or below-average task-scores, relative to their year, while the under-confident and over-confident groups were unchanged (**Table 21**). These groups then broadly matched the idea underlying the conceptual (above-average/below-average) accuracy/bias groups that were considered in PISA 2006, but directly considered self-reflective accuracy/bias via the difference-score indicator.

Various differences in attitudes and beliefs were observed across the groups; means for the groups are presented in **Table 21** and differences across groups are presented in **Table 22**. For example, under-confident students reported similar interest, utility, personal, and cost values and norms/influences from their parents, to those with accurately-low beliefs, despite reporting higher current science grades. Those with accurately-high beliefs reported the highest attitudes and intentions (above-average, relative to their academic year), significantly higher than over-confident students. The students' reported background characteristics also varied across the groups, for example with above-average proportions of boys being present in the accurately-high group, roughly average proportions of boys in the over-confident groups, and similar and below-average proportions of boys in the under-confident and accurately-low groups.

Students' science intentions differed across all potential pairs of groups, except for the under-confident and over-confident groups. Fundamentally, however, only a small amount of variance in students' science intentions (7.6%) could be attributed to the differences across the groups. For illustration, considering the students' science intentions on the original agreement-scales from (1) 'strongly disagree' to (6) 'strongly agree', accurately-high students essentially 'slightly agreed' (M = 4.16, SD = 1.32) , under-confident and over-confident students similarly reported around the mid-point (under-confident M = 3.33, SD = 1.49, and over-confident M = 3.50, SD = 1.49, between 'slightly agree' and 'slightly disagree'), while accurately-low students essentially 'slightly disagreed' (M = 3.00, SD = 1.36).

Table 21: England 2014/2015 (Years 9, 10, 11): difference-score accuracy/bias groups with accurately-low/accurately-high groups

	Under-confident (U)		Accurately-low (L)		Accurately-high (H)		Over-confident (O)	
Item/factor (per year *z*-scores)	M	SD	M	SD	M	SD	M	SD
Science intentions	-.10	1.00	-.34	.91	.47	.91	.02	1.00
Task-score	.67	.69	-.62	.49	.82	.47	-.80	.87
Task-confidence	-.42	.83	-.75	.62	.92	.64	.32	.94
Task-conf. accuracy/bias (diff.-score)	-1.12	.52	.01	.27	-.06	.27	1.17	.67
Task sensitivity	-.38	.90	-.62	.95	.77	.44	.30	.97
Task specificity	.52	.83	.52	.69	-.56	.94	-.50	.92
Task simple-matching	-.18	.94	.01	.85	.62	.71	-.26	1.13
Task-conf. accuracy/bias (regression-r.)	-.89	.71	-.50	.48	.58	.51	.86	.80
Self-concept accuracy/bias (regression-r.)	-.30	.97	-.21	.97	.37	.81	.18	1.04
Gender (1=boy)	-.23	1.01	-.22	1.01	.43	.82	.07	.99
Ethnicity (White)	.01	1.00	.27	.87	-.23	1.02	-.02	1.02
Ethnicity (Black)	.01	1.06	-.04	.88	.07	1.21	-.04	.82
Ethnicity (East-Asian)	-.05	.79	-.05	.79	.17	1.47	-.03	.90
Ethnicity (South-Asian/Indian)	.01	.99	-.26	.77	.14	1.06	.05	1.05
Ethnicity (mixed)	-.01	.98	-.02	.95	.04	1.09	-.01	.99
Ethnicity (other)	-.05	.71	.02	1.12	.05	1.17	.00	1.05
Books at home	.12	.95	-.27	.98	.38	.89	-.20	1.04
Parent(s) in science (1=yes)	.02	1.02	-.11	.91	.08	1.06	.00	.99
Mothers' education	.01	1.00	-.31	.86	.37	1.01	-.04	1.01
Fathers' education	-.02	1.01	-.34	.86	.37	1.01	.00	.99
Current grade	.08	.96	-.48	.85	.61	.88	-.17	1.00
Self-concept	-.22	.96	-.40	.90	.63	.77	.06	1.03
Self-efficacy	-.03	.99	-.45	.90	.70	.69	-.13	1.01
Interest value	-.21	1.02	-.32	.92	.57	.81	.04	1.00
Utility value	-.17	1.04	-.26	.96	.48	.81	.02	.99
Personal value	-.22	.99	-.28	.91	.48	.89	.07	1.02
Cost value (absence of)	.00	.99	.00	1.01	.16	.96	-.10	1.02

Item/factor (per year *z*-scores)	Under-confident (U)		Accurately-low (L)		Accurately-high (H)		Over-confident (O)	
	M	SD	M	SD	M	SD	M	SD
Orientation: mastery	-.03	1.01	-.26	.98	.42	.78	-.08	1.07
Orientation: performance	-.10	1.00	-.24	1.00	.40	.85	-.02	1.03
Perceived control	-.09	.98	-.13	1.04	.41	.71	-.09	1.10
Perceived control (exams)	-.10	.98	-.26	.98	.53	.82	-.08	1.03
Study strategy: self-regulation	-.21	1.01	-.21	.92	.41	.90	.08	1.03
Study strategy: control	-.11	1.03	-.17	.94	.30	.88	.03	1.05
Study strategy: memorisation	-.11	.98	-.20	.97	.27	.92	.06	1.07
Study strategy: elaboration	-.26	.92	-.15	.99	.33	.87	.13	1.09
Anxiety (absence of)	-.19	.97	-.36	.98	.58	.77	.03	1.01
Mastery norms (good grade)	-.02	.98	-.07	.99	.19	.87	-.05	1.09
Subject-comparisons	-.16	.99	-.27	1.02	.47	.85	.02	.98
Peer-comparisons	-.21	1.04	-.22	.99	.48	.78	.02	.99
Social persuasions (praise)	-.20	.98	-.27	.98	.48	.81	.06	1.04
Vicarious experiences	-.12	1.01	-.18	.98	.28	.89	.06	1.05
Norms/influence (friends)	-.02	.91	-.06	.96	.15	.94	-.04	1.17
Norms/influence (parents)	-.11	1.01	-.21	.97	.43	.83	-.04	1.04
Teacher perceptions	-.15	.96	-.23	1.00	.38	.84	.06	1.07
Teacher/school careers/events	-.14	.92	-.22	.98	.22	.92	.14	1.10
Group size (N, %)	441	29.4%	318	21.2%	321	21.4%	422	28.1%

*Notes: Missing responses were estimated by expectation-maximisation. Means (M) and standard deviations (SD) are reported. Science intentions and their key predictors (identified in **Section 7**) are shaded in grey.*

Table 22: England 2014/2015 (Years 9, 10, 11): difference-score accuracy/bias groups with accurately-low/accurately-high groups, group differences

	Overall group difference		Paired group differences, sig. (*p*)					
Item/factor	Sig. (*p*)	Size (η^2)	U-L	U-H	U-O	L-H	L-O	H-O
Science intentions	**<.001**	.076	**.005**	**<.001**	.385	**<.001**	**<.001**	**<.001**
Task-score	**<.001**	.545	**<.001**	**.021**	**<.001**	**<.001**	**.001**	**<.001**
Task-confidence	**<.001**	.381	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Task-conf. accuracy/bias (diff.-score)	**<.001**	.761	**<.001**	**<.001**	**<.001**	.596	**<.001**	**<.001**
Task sensitivity	**<.001**	.275	**.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Task specificity	**<.001**	.273	1.000	**<.001**	**<.001**	**<.001**	**<.001**	1.000
Task simple-matching	**<.001**	.112	**.037**	**<.001**	1.000	**<.001**	**.001**	**<.001**
Task-conf. accuracy/bias (regression-r.)	**<.001**	.566	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-concept accuracy/bias (regression-r.)	**<.001**	.078	1.000	**<.001**	**<.001**	**<.001**	**<.001**	.102
Gender (1=boy)	**<.001**	.066	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Ethnicity (White)	**<.001**	.026	**.002**	**.007**	1.000	**<.001**	**.001**	**.026**
Ethnicity (Black)	.400	.002	1.000	1.000	1.000	.993	1.000	.729
Ethnicity (East-Asian)	**.011**	.007	1.000	**.020**	1.000	**.033**	1.000	.058
Ethnicity (South-Asian/Indian)	**<.001**	.019	**.001**	.521	1.000	**<.001**	**<.001**	1.000
Ethnicity (mixed)	.865	<.001	1.000	1.000	1.000	1.000	1.000	1.000
Ethnicity (other)	.596	.001	1.000	1.000	1.000	1.000	1.000	1.000
Books at home	**<.001**	.061	**<.001**	**.002**	**<.001**	**<.001**	1.000	**<.001**
Parent(s) in science (1=yes)	.113	.004	.403	1.000	1.000	.111	.910	1.000
Mothers' education	**<.001**	.051	**<.001**	**<.001**	1.000	**<.001**	**.001**	**<.001**
Fathers' education	**<.001**	.054	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**
Current grade	**<.001**	.138	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-concept	**<.001**	.135	**.046**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-efficacy	**<.001**	.152	**<.001**	**<.001**	.610	**<.001**	**<.001**	**<.001**
Interest value	**<.001**	.105	.625	**<.001**	**.001**	**<.001**	**<.001**	**<.001**
Utility value	**<.001**	.072	1.000	**<.001**	**.020**	**<.001**	**<.001**	**<.001**
Personal value	**<.001**	.081	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Cost value (absence of)	**.005**	.008	1.000	.197	.690	.236	1.000	**.002**

Item/factor	Overall group difference		Paired group differences, sig. (p)					
	Sig. (p)	Size (η^2)	U-L	U-H	U-O	L-H	L-O	H-O
Orientation: mastery	**<.001**	.054	**.007**	**<.001**	1.000	**<.001**	.068	**<.001**
Orientation: performance	**<.001**	.049	.363	**<.001**	1.000	**<.001**	**.017**	**<.001**
Perceived control	**<.001**	.045	1.000	**<.001**	1.000	**<.001**	1.000	**<.001**
Perceived control (exams)	**<.001**	.078	.155	**<.001**	1.000	**<.001**	.083	**<.001**
Study strategy: self-regulation	**<.001**	.060	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Study strategy: control	**<.001**	.029	1.000	**<.001**	.229	**<.001**	**.037**	**.002**
Study strategy: memorisation	**<.001**	.028	1.000	**<.001**	.081	**<.001**	**.003**	**.031**
Study strategy: elaboration	**<.001**	.052	.916	**<.001**	**<.001**	**<.001**	**.001**	**.024**
Anxiety (absence of)	**<.001**	.110	.093	**<.001**	**.003**	**<.001**	**<.001**	**<.001**
Mastery norms (good grade)	**.002**	.010	1.000	**.022**	1.000	**.006**	1.000	**.007**
Subject-comparisons	**<.001**	.069	.555	**<.001**	.054	**<.001**	**<.001**	**<.001**
Peer-comparisons	**<.001**	.071	1.000	**<.001**	.004	**<.001**	**.006**	**<.001**
Social persuasions (praise)	**<.001**	.078	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Vicarious experiences	**<.001**	.029	1.000	**<.001**	**.050**	**<.001**	**.007**	**.014**
Norms/influence (friends)	**.020**	.007	1.000	.090	1.000	**.039**	1.000	**.046**
Norms/influence (parents)	**<.001**	.053	1.000	**<.001**	1.000	**<.001**	.103	**<.001**
Teacher perceptions	**<.001**	.050	1.000	**<.001**	**.008**	**<.001**	**<.001**	**<.001**
Teacher/school careers/events	**<.001**	.031	1.000	**<.001**	**<.001**	**<.001**	**<.001**	1.000

Notes: Missing responses were estimated by expectation-maximisation. Significant p-values (p < .05) are highlighted in bold. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey.*

Section 8.2.2: Accuracy/bias clusters

In order to identify natural clusters within the data, latent-profile analysis was undertaken using the students' task-confidence and task-scores, both standardised as within-year *z*-scores (**Section 6.3**). Any clusters would then exhibit distinct magnitudes of task-confidence and task-scores, and hence likely exhibit distinct magnitudes of accuracy/bias.

Preliminary analysis produced broadly similar results (i.e. cluster profiles and proportions) for individual academic years considered separately, especially for lower numbers of modelled clusters (e.g. four clusters). However, manually combining/linking these various cluster assignments across the different years was not necessarily feasible, especially when modelling higher numbers of clusters (i.e. in situations where any clusters did not obviously match). Additionally, from the preliminary analysis, it was not necessarily clear whether differences in emerging cluster profiles and proportions across different academic years related to actual differences or followed from unreliability due to lower numbers of students being considered in Year 10 and Year 11 compared to Year 9.

Accordingly, for operational efficiency, it was necessary to assume that cluster profiles and proportions were similar within each academic year (an assumption which was indeed broadly supported by the preliminary results). Latent-profile analysis was then undertaken across Years 9-11 considered together, and used standardised indicators (i.e. within-year *z*-scores) to account for potential differences in means per year.

Essentially, the approach identified discrete clusters of students with particular magnitudes of task-score and task-confidence (in any combination that might emerge, and hence with any potential degree of accuracy/bias), maximising similarities within-clusters and differences across-clusters (Collins & Lanza, 2010; Vermunt & Magidson, 2013). For example, the approach might identify a discrete cluster of students all with above-average task-scores and below-average task-confidence (relative to other students within their academic year), if that was a discrete cluster of students that naturally-occurred within the data.

Identifying the number of clusters to consider

Modelling one to ten clusters highlighted less relative improvements in the model information criteria from five clusters onwards (**Table 23**); lower information criteria reflected better fit to the data. However, given the varying cluster sizes, predictive modelling (necessary to address the subsequent research question in **Section 9**) was only feasible for each of the clusters from the four-cluster solution: the smallest cluster from the five-cluster solution only included 69 students.

Similarities in clusters could be observed across the different solutions, highlighting that some distinct clusters were identified regardless of the number of clusters being modelled. For example, three of the four clusters from the four-cluster solution remained as relatively discrete clusters in the five-cluster solution. Similarly, three clusters remained relatively discrete across the five-cluster and the six-cluster solutions. Fundamentally, considering four clusters (rather than five clusters) appeared to be a feasible compromise between the information criteria and potential comparability with PISA 2006 and the other grouping approaches, especially given the similar clusters across the four-cluster and five-cluster solutions (see **Appendix 9** for the five-cluster details). Further research with larger samples may benefit from considering higher numbers of clusters.

Accuracy/bias clusters from latent-profile analysis (four clusters)

The cluster profiles are summarised in **Table 24**, and differences across the clusters are summarised in **Table 25**. Tabulating the clusters against the difference-score under-confident, accurately-low, accurately-high, and over-confident groups highlighted that cluster A, the largest cluster, was mainly composed of ‘under-confident’ students (48.3%) then ‘accurately-high’ students (34.5%). Cluster B was mainly composed of ‘accurately-low’ students (44.7%) then ‘over-confident’ students (35.6%). Cluster C was mainly composed of ‘accurately-high’ students (76.2%), while cluster D was mainly composed of ‘over-confident’ students (83.2%).

Students in cluster A (42.3% of the modelled students) exhibited moderately above-average task-scores and task-confidence, although with a small bias towards under-confidence. Students in cluster B (41.5%) exhibited below-average task-scores and task-confidence, and were essentially accurate. Students in cluster C (8.7%) exhibited highly above-average task-scores and task-confidence, and were again essentially accurate. Students in cluster D (7.5%) exhibited highly below-average task-scores and moderately below-average task-confidence, and hence over-confidence.

Considered generally, the profiles were remarkably similar to the conceptual idea of four groups, broadly covering under-confident (cluster A), accurately-low (cluster B), accurately-high (cluster C), and over-confident (cluster D) students. The clusters nevertheless exhibited distinct magnitudes of their task-scores and task-confidences; for example, students in cluster A exhibited moderately above-average task-scores and lower but still above-average task-confidence, rather than reflecting a conceptual idea of 'above-average task-scores but below-average task-confidence'.

Students in the different clusters had different magnitudes of (difference-score) accuracy/bias across all potential pairs, except for clusters B and C. For the other indicators of accuracy, the various pairs of clusters differed in task sensitivity and specificity, except for clusters B and D; similarly, the clusters differed in task simple-matching except for clusters A and D and for clusters B and D. Fundamentally, students in cluster C exhibited the highest sensitivity (knowing when they were right) and simple-matching (knowing when they were right and also when they were wrong), but the lowest specificity (knowing when they were wrong). Students in cluster C gained very high scores (on average, scoring 94% correct) and hence 'ceiling effects' may have been relevant (i.e. fewer extremely difficult tasks where they might have been wrong, hence they may have had fewer opportunities to recognise that they were wrong). Students in cluster C also exhibited the highest degree of relative over-confidence via the regression-residual indicators, highlighting that relative accuracy/bias does not necessarily reflect absolute accuracy/bias (which potentially offers insight into the PISA 2006 results that were only able to consider a relative regression-residual indicator of accuracy/bias).

Additionally, the results for cluster D may be potentially harder to infer from, given that these students gained extremely low scores (on average, scoring 2% correct); these students appeared to attempt some tasks in some manner, although this may have reflected disengagement, guessing, or other potential tendencies.

The students' reported science grades, self-concept beliefs, and self-efficacy beliefs broadly followed the pattern of magnitudes seen for their task-scores and task-confidence. The students' various attitudes followed a similar pattern, where students in cluster C exhibited the highest, above-average attitudes and beliefs, while students in cluster A exhibited slightly above-average or average attitudes, while students in clusters B and D exhibited slightly or moderately below-average attitudes. For the majority of the attitudinal measures, students in clusters B and D did not significantly differ in their beliefs. Students in cluster C also reported the highest (above-average) tendencies towards self-regulated studying strategies, and control, memorisation, and elaboration strategies; all of the various pairs of clusters significantly differed for these measures, except for clusters B and D.

The students' background and other reported characteristics also varied across the clusters, for example with above-average proportions of boys being present in cluster C (with the highest attitudes and beliefs), who also reported the highest levels of books at home, parental education, and parents working within science.

The students' science intentions differed across the clusters, on average, and across all potential pairs of clusters except that no difference was apparent between clusters B and D. Students in cluster C expressed the highest, above-average intentions (relative to students in their respective year). However, only a small amount of variance in the students' science intentions (13.5%) could be attributed to the differences across the clusters. When considered on the unstandardized agreement-scale (from 1 to 6), the average response for cluster C was closer to agreement (M = 4.76, SD = 1.19), while students in cluster A averaged between slight agreement and slight disagreement (M = 3.77, SD = 1.41, around but just above a neutral mid-point of 3.50), while students in clusters B and D averaged around slight disagreement (cluster B with M = 2.98, SD = 1.39; cluster D with M = 3.16, SD = 1.31).

Table 23: England 2014/2015 (Years 9, 10, 11): task-level accuracy/bias clusters, model information criteria

Clusters	BIC	AIC	SA-BIC
1	8528.99	8507.73	8516.29
2	8000.86	7953.03	7972.27
3	7926.07	7851.67	7881.60
4	7883.72	7782.75	7823.36
5	7829.02	7701.47	7752.78
6	7838.08	7683.96	7745.95
7	7797.90	7617.21	7689.89
8	7804.30	7597.03	7680.41
9	7828.43	7594.59	7688.65
10	7829.93	7569.52	7674.27

Notes: The Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and sample-size adjusted Bayesian Information Criterion (SA-BIC) from the respective solutions for the numbers of clusters are reported. Lower information criteria reflected better fit to the data.

Table 24: England 2014/2015 (Years 9, 10, 11): task-level accuracy/bias clusters (four clusters)

	Cluster A		Cluster B		Cluster C		Cluster D	
Item/factor (per year *z*-scores)	M	SD	M	SD	M	SD	M	SD
Science intentions	.21	.95	-.35	.94	.86	.84	-.21	.87
Task-score	.70	.47	-.62	.45	1.36	.23	-1.90	.17
Task-confidence	.33	.72	-.55	.75	1.69	.30	-.69	1.05
Task-conf. accuracy/bias (diff.-score)	-.47	.88	.18	.85	.04	.38	1.47	1.02
Task sensitivity	.32	.76	-.46	1.08	1.08	.15	-.42	.74
Task specificity	-.14	1.04	.29	.85	-.89	.73	.18	1.07
Task simple-matching	.07	.84	-.29	.88	1.22	.31	-.11	1.63
Task-conf. accuracy/bias (regression-r.)	-.04	.92	-.27	.92	1.15	.39	.36	1.27
Self-concept accuracy/bias (regression-r.)	.04	.97	-.14	1.01	.63	.72	-.34	1.03
Gender (1=boy)	.13	.97	-.25	1.01	.63	.62	-.13	1.03
Ethnicity (White)	-.10	1.02	.27	.88	-.56	.92	-.23	1.11
Ethnicity (Black)	.00	1.02	-.04	.88	.12	1.31	.11	1.12
Ethnicity (East-Asian)	-.01	.95	-.08	.67	.52	2.08	-.07	.81
Ethnicity (South-Asian/Indian)	.11	1.07	-.24	.76	.36	1.05	.21	1.24
Ethnicity (mixed)	.04	1.08	-.05	.89	-.03	.93	.09	1.18
Ethnicity (other)	-.03	.78	.00	1.05	.18	1.61	-.01	1.02
Books at home	.26	.90	-.27	.98	.62	.76	-.67	1.01
Parent(s) in science (1=yes)	.04	1.02	-.11	.91	.35	1.19	-.06	.97
Mothers' education	.19	1.01	-.31	.88	.73	.92	-.14	1.02
Fathers' education	.15	1.00	-.33	.88	.90	.85	-.07	.97
Current grade	.33	.87	-.46	.86	1.19	.59	-.64	.90
Self-concept	.20	.88	-.34	.92	1.13	.70	-.55	1.10
Self-efficacy	.34	.82	-.47	.92	1.13	.42	-.49	1.02
Interest value	.23	.86	-.34	.99	.98	.65	-.51	.97
Utility value	.19	.90	-.31	1.01	.81	.65	-.26	1.03
Personal value	.15	.93	-.31	.96	.97	.75	-.24	1.00
Cost value (absence of)	-.02	.96	.02	1.01	.31	1.05	-.24	1.02

	Cluster A		Cluster B		Cluster C		Cluster D	
Item/factor (per year *z*-scores)	M	SD	M	SD	M	SD	M	SD
Orientation: mastery	.21	.85	-.27	1.05	.75	.66	-.46	1.12
Orientation: performance	.14	.94	-.25	1.02	.73	.76	-.22	.94
Perceived control	.24	.73	-.21	1.09	.54	.90	-.68	1.25
Perceived control (exams)	.18	.95	-.23	.98	.73	.78	-.51	.96
Study strategy: self-regulation	.15	.92	-.26	.97	.85	.90	-.37	1.05
Study strategy: control	.12	.91	-.19	1.02	.64	.84	-.31	1.11
Study strategy: memorisation	.15	.88	-.20	1.05	.48	.98	-.31	1.08
Study strategy: elaboration	.08	.90	-.18	1.02	.64	.89	-.22	1.16
Anxiety (absence of)	.20	.90	-.33	.99	.93	.63	-.36	1.04
Mastery norms (good grade)	.07	.92	-.14	1.04	.53	.78	-.11	1.14
Subject-comparisons	.13	.95	-.26	1.00	.89	.58	-.29	.96
Peer-comparisons	.12	.94	-.24	.99	.79	.71	-.27	1.09
Social persuasions (praise)	.17	.90	-.30	1.01	.88	.79	-.24	.92
Vicarious experiences	.09	.93	-.15	1.03	.55	.89	-.29	1.09
Norms/influence (friends)	.03	.91	-.07	1.04	.27	1.05	-.11	1.20
Norms/influence (parents)	.16	.95	-.25	1.01	.69	.67	-.24	1.06
Teacher perceptions	.15	.86	-.20	1.05	.62	.85	-.39	1.19
Teacher/school careers/events	.06	.95	-.18	1.02	.52	.88	.03	1.08
Cluster size (N, %)	635	42.3%	624	41.5%	130	8.7%	113	7.5%

Notes: Missing responses were estimated by expectation-maximisation. Means (M) and standard deviations (SD) are reported. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey.*

Table 25: England 2014/2015 (Years 9, 10, 11): task-level accuracy/bias clusters (four clusters), cluster differences

	Overall group difference		Paired group differences, sig. (*p*)					
Item/factor	Sig. (*p*)	Size (η^2)	A-B	A-C	A-D	B-C	B-D	C-D
Science intentions	**<.001**	.135	**<.001**	**<.001**	**<.001**	**<.001**	.823	**<.001**
Task-score	**<.001**	.814	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Task-confidence	**<.001**	.457	**<.001**	**<.001**	**<.001**	**<.001**	.383	**<.001**
Task-conf. accuracy/bias (diff.-score)	**<.001**	.273	**<.001**	**<.001**	**<.001**	.454	**<.001**	**<.001**
Task sensitivity	**<.001**	.242	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Task specificity	**<.001**	.114	**<.001**	**<.001**	**.006**	**<.001**	1.000	**<.001**
Task simple-matching	**<.001**	.168	**<.001**	**<.001**	.325	**<.001**	.306	**<.001**
Task-conf. accuracy/bias (regression-r.)	**<.001**	.156	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-concept accuracy/bias (regression-r.)	**<.001**	.054	**.012**	**<.001**	**.012**	**<.001**	.666	**<.001**
Gender (1=boy)	**<.001**	.068	**<.001**	**<.001**	.063	**<.001**	1.000	**<.001**
Ethnicity (White)	**<.001**	.066	**<.001**	**<.001**	1.000	**<.001**	**<.001**	.052
Ethnicity (Black)	.250	.003	1.000	1.000	1.000	.599	.935	1.000
Ethnicity (East-Asian)	**<.001**	.026	1.000	**<.001**	1.000	**<.001**	1.000	**<.001**
Ethnicity (South-Asian/Indian)	**<.001**	.046	**<.001**	.060	1.000	**<.001**	**<.001**	1.000
Ethnicity (mixed)	.372	.002	.885	1.000	1.000	1.000	1.000	1.000
Ethnicity (other)	.203	.003	1.000	.196	1.000	.375	1.000	.927
Books at home	**<.001**	.127	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Parent(s) in science (1=yes)	**<.001**	.016	.064	**.009**	1.000	**<.001**	1.000	**.012**
Mothers' education	**<.001**	.103	**<.001**	**<.001**	**.004**	**<.001**	.448	**<.001**
Fathers' education	**<.001**	.125	**<.001**	**<.001**	.124	**<.001**	**.040**	**<.001**
Current grade	**<.001**	.288	**<.001**	**<.001**	**<.001**	**<.001**	.197	**<.001**
Self-concept	**<.001**	.198	**<.001**	**<.001**	**<.001**	**<.001**	.126	**<.001**
Self-efficacy	**<.001**	.269	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Interest value	**<.001**	.173	**<.001**	**<.001**	**<.001**	**<.001**	.454	**<.001**
Utility value	**<.001**	.118	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Personal value	**<.001**	.134	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Cost value (absence of)	**<.001**	.013	1.000	**.003**	.197	**.013**	.069	**<.001**

Item/factor	Overall group difference Sig. (p)	Size (η^2)	Paired group differences, sig. (p) A-B	A-C	A-D	B-C	B-D	C-D
Orientation: mastery	**<.001**	.114	**<.001**	**<.001**	**<.001**	**<.001**	.280	**<.001**
Orientation: performance	**<.001**	.082	**<.001**	**<.001**	**.002**	**<.001**	1.000	**<.001**
Perceived control	**<.001**	.101	**<.001**	**.007**	**<.001**	**<.001**	**<.001**	**<.001**
Perceived control (exams)	**<.001**	.101	**<.001**	**<.001**	**<.001**	**<.001**	**.025**	**<.001**
Study strategy: self-regulation	**<.001**	.109	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Study strategy: control	**<.001**	.064	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Study strategy: memorisation	**<.001**	.052	**<.001**	**.003**	**<.001**	**<.001**	1.000	**<.001**
Study strategy: elaboration	**<.001**	.054	**<.001**	**<.001**	**.014**	**<.001**	1.000	**<.001**
Anxiety (absence of)	**<.001**	.145	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Mastery norms (good grade)	**<.001**	.036	**.001**	**<.001**	.525	**<.001**	1.000	**<.001**
Subject-comparisons	**<.001**	.111	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Peer-comparisons	**<.001**	.090	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Social persuasions (praise)	**<.001**	.122	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Vicarious experiences	**<.001**	.045	**<.001**	**<.001**	**.001**	**<.001**	1.000	**<.001**
Norms/influence (friends)	**.003**	.009	.583	.077	1.000	**.003**	1.000	**.021**
Norms/influence (parents)	**<.001**	.082	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Teacher perceptions	**<.001**	.071	**<.001**	**<.001**	**<.001**	**<.001**	.302	**<.001**
Teacher/school careers/events	**<.001**	.038	**<.001**	**<.001**	1.000	**<.001**	.208	**.001**

Notes: Missing responses were estimated by expectation-maximisation. Significant p-values (p < .05) are highlighted in bold. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey.*

Section 8.2.3: Intention/attitude clusters

As with PISA 2006, latent-profile analysis was also applied to form clusters based on students' science intentions and key predictors. Any differences in task-level accuracy/bias could then be considered across the clusters. This also offered the potential to quantify the proportion of students who might or might not be expected to progress further within science, given their expressed intentions and attitudes. Accordingly, latent-profile analysis was used to reveal clusters of students given their science intentions, self-efficacy, interest, perceived utility, personal value, and norms/influences from parents (i.e. the key predictors and outcome from **Section 7)**. These items/factors were standardised (as *z*-scores per academic year) and Years 9-11 were considered together.

Modelling one to ten clusters highlighted less relative improvements in the model information criteria from four or five clusters onwards (**Table 26**). Less relative change to the cluster compositions was also observed from four-clusters to five-clusters and subsequently (compared to, for example, the change in cluster compositions from three-clusters to four-clusters). It was again plausible to primarily consider the four-cluster solution (see **Appendix 9** for the five-cluster solution). For these intention/attitude clusters, there appeared to be no clear correspondences with the difference-score accuracy/bias groups or with the accuracy/bias clusters.

Considering four intention/attitude clusters, the cluster profiles and differences are respectively presented in **Table 27** and **Table 28**. The four-cluster solution essentially contained one small cluster of students (D, 12.9% of considered students) who expressed highly below-average intentions and attitudes (relative to other students within their respective academic year), one cluster moderately below-average (A, 37.3%), one cluster moderately above-average (B, 33.6%), and another cluster highly above-average (C, 12.9%). The students' reported science grades, task-scores, task-confidence, and also other attitudes, similarly followed this pattern of magnitudes.

Considered against the underlying agreement-scale (from 1 to 6), students in cluster D disagreed with intending to studying further in science

(M = 1.77, SD = .96), students in cluster A responded with slight disagreement (M = 2.64, SD = 1.08), students in cluster B responded with slight agreement (M = 4.06, SD = 1.02), and students in cluster C strongly agreed to study science further (M = 5.56, SD = .53). A large proportion of variance in science intentions (61.1%) could be attributed to the difference across the clusters, which was perhaps less informative given that the students were directly clustered according to their science intentions (and key predicting factors).

The students' (difference-score) accuracy/bias had less variation across the clusters, with differences only observable between clusters A and D and clusters C and D. Essentially, those in the cluster with the lowest intentions and attitudes (cluster D), exhibited a higher degree of under-confidence than some other clusters. The indicators of task sensitivity and specificity differed across all pairs of clusters, while the simple-matching indicator of accuracy differed across the clusters on average, but only across some pairs of clusters (and not between A and D, and B and D). Essentially, those in the cluster with the highest intentions and attitudes (cluster C) exhibited the highest task-level accuracy considered as sensitivity (measuring whether they accurately knew when they had right answers) and simple-matching (measuring whether they accurately knew when they had right and wrong answers), but the lowest specificity (measuring whether they accurately knew when they had wrong answers).

Table 26: England 2014/2015 (Years 9, 10, 11): intention/attitude clusters, model information criteria

Clusters	BIC	AIC	SA-BIC
1	19646.21	19585.54	19608.09
2	17018.05	16891.66	16938.64
3	16088.35	15896.24	15967.65
4	15723.70	15465.88	15561.71
5	15584.95	15261.41	15381.67
6	15527.77	15138.51	15283.20
7	15444.08	14989.10	15158.21
8	15405.06	14884.37	15077.90
9	15399.36	14812.94	15030.91
10	15385.97	14733.83	14976.22

Notes: The Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and sample-size adjusted Bayesian Information Criterion (SA-BIC) from the respective solutions for the numbers of clusters are reported. Lower information criteria reflected better fit to the data.

Table 27: England 2014/2015 (Years 9, 10, 11): intention/attitude clusters (four clusters)

	Cluster A		Cluster B		Cluster C		Cluster D	
Item/factor (per year *z*-scores)	M	SD	M	SD	M	SD	M	SD
Science intentions	-.58	.71	.38	.69	1.42	.37	-1.18	.65
Task-score	-.28	.84	.21	.86	.65	.79	-.42	.88
Task-confidence	-.31	.82	.17	.86	.93	.85	-.75	.84
Task-conf. accuracy/bias (diff.-score)	.02	.90	-.07	.86	.11	.91	-.22	.97
Task sensitivity	-.21	1.00	.14	.91	.65	.67	-.56	.98
Task specificity	.21	.87	-.11	1.01	-.70	.96	.59	.76
Task simple-matching	-.14	.90	.02	.85	.34	.88	-.07	.85
Task-conf. accuracy/bias (regression-r.)	-.20	.87	.08	.86	.68	.92	-.61	.93
Self-concept accuracy/bias (regression-r.)	-.26	.85	.22	.84	.69	.90	-.77	.99
Gender (1=boy)	-.19	1.01	.11	.98	.22	.95	-.39	1.01
Ethnicity (White)	.24	.86	-.08	.99	-.26	1.05	.35	.79
Ethnicity (Black)	-.08	.70	-.06	.84	-.18	.05	-.01	1.04
Ethnicity (East-Asian)	-.03	.87	-.02	.93	.14	1.43	-.05	.79
Ethnicity (South-Asian/Indian)	-.21	.79	.14	1.05	.30	1.13	-.33	.65
Ethnicity (mixed)	-.01	.99	-.03	.93	.02	1.05	-.02	.96
Ethnicity (other)	-.06	.62	.00	1.01	.04	1.10	-.10	.02
Books at home	-.11	.99	.13	.93	.41	.93	-.24	1.03
Parent(s) in science (1=yes)	-.03	.97	.00	1.00	.21	1.14	-.23	.79
Mothers’ education	-.16	.96	.10	.99	.32	1.02	-.36	.87
Fathers’ education	-.21	.92	.07	.99	.43	1.02	-.46	.88
Current grade	-.23	.90	.26	.92	.75	.84	-.65	.97
Self-concept	-.32	.78	.34	.75	.99	.87	-.96	.87
Self-efficacy	-.25	.89	.30	.85	.86	.57	-.78	1.02
Interest value	-.36	.66	.48	.52	1.27	.37	-1.49	.85
Utility value	-.42	.56	.53	.44	1.32	.25	-1.65	.62
Personal value	-.45	.63	.41	.60	1.40	.42	-1.51	.39
Cost value (absence of)	.00	.90	-.09	1.00	-.09	1.27	.47	1.18

	Cluster A		Cluster B		Cluster C		Cluster D	
Item/factor (per year *z*-scores)	M	SD	M	SD	M	SD	M	SD
Orientation: mastery	-.21	.93	.24	.78	.88	.60	-.85	1.29
Orientation: performance	-.21	.94	.23	.85	.75	.80	-.67	1.14
Perceived control	-.14	.91	.28	.78	.86	.59	-.85	1.36
Perceived control (exams)	-.12	.91	.22	.95	.73	1.05	-.48	1.11
Study strategy: self-regulation	-.24	.81	.23	.84	1.01	.87	-1.08	1.00
Study strategy: control	-.22	.93	.25	.74	.98	.76	-.87	1.25
Study strategy: memorisation	-.22	.94	.22	.83	.76	.93	-.99	1.11
Study strategy: elaboration	-.27	.86	.24	.90	.85	.91	-1.06	1.02
Anxiety (absence of)	-.26	.92	.21	.87	.67	.96	-.59	1.08
Mastery norms (good grade)	-.09	.99	.00	1.02	.22	.93	-.15	1.12
Subject-comparisons	-.27	.94	.30	.86	.62	.92	-.64	1.07
Peer-comparisons	-.21	.92	.21	.87	.55	.99	-.55	1.15
Social persuasions (praise)	-.33	.85	.34	.79	.93	.87	-.82	.98
Vicarious experiences	-.10	.95	.28	.81	.52	1.01	-.76	1.18
Norms/influence (friends)	-.14	.92	.19	.93	.45	1.04	-.38	1.19
Norms/influence (parents)	-.38	.80	.51	.57	.86	.80	-1.12	.96
Teacher perceptions	-.25	.91	.27	.81	.92	.74	-.92	1.15
Teacher/school careers/events	-.24	.91	.21	.95	.51	1.14	-.81	.95
Cluster size (N, %)	432	37.3%	389	33.6%	188	16.2%	150	12.9%

*Notes: Missing responses were estimated by expectation-maximisation. Means (M) and standard deviations (SD) are reported. Science intentions and their key predictors (identified in **Section 7**) are shaded in grey; the clusters were formed on these factors.*

Table 28: England 2014/2015 (Years 9, 10, 11): intention/attitude clusters (four clusters), cluster differences

	Overall group difference		Paired group differences, sig. (*p*)					
Item/factor	Sig. (*p*)	Size (η^2)	A-B	A-C	A-D	B-C	B-D	C-D
Science intentions	**<.001**	.616	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Task-score	**<.001**	.160	**<.001**	**<.001**	.420	**<.001**	**<.001**	**<.001**
Task-confidence	**<.001**	.268	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Task-conf. accuracy/bias (diff.-score)	**.004**	.011	.869	1.000	**.036**	.118	.593	**.005**
Task sensitivity	**<.001**	.134	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Task specificity	**<.001**	.147	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Task simple-matching	**<.001**	.035	**.045**	**<.001**	1.000	**<.001**	1.000	**<.001**
Task-conf. accuracy/bias (regression-r.)	**<.001**	.153	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-concept accuracy/bias (regression-r.)	**<.001**	.205	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Gender (1=boy)	**<.001**	.042	**<.001**	**<.001**	.211	1.000	**<.001**	**<.001**
Ethnicity (White)	**<.001**	.051	**<.001**	**<.001**	1.000	.192	**<.001**	**<.001**
Ethnicity (Black)	.192	.004	1.000	.781	1.000	.417	1.000	.270
Ethnicity (East-Asian)	.208	.004	1.000	.307	1.000	.469	1.000	.489
Ethnicity (South-Asian/Indian)	**<.001**	.057	**<.001**	**<.001**	1.000	.323	**<.001**	**<.001**
Ethnicity (mixed)	.943	<.001	1.000	1.000	1.000	1.000	1.000	1.000
Ethnicity (other)	.326	.003	1.000	1.000	1.000	1.000	1.000	.752
Books at home	**<.001**	.044	**.002**	**<.001**	.937	.008	**<.001**	**<.001**
Parent(s) in science (1=yes)	**.001**	.015	1.000	**.027**	.274	.104	.115	**<.001**
Mothers' education	**<.001**	.046	**.001**	**<.001**	.191	.080	**<.001**	**<.001**
Fathers' education	**<.001**	.075	**<.001**	**<.001**	**.036**	**<.001**	**<.001**	**<.001**
Current grade	**<.001**	.184	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-concept	**<.001**	.358	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-efficacy	**<.001**	.256	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Interest value	**<.001**	.645	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Utility value	**<.001**	.769	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Personal value	**<.001**	.699	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Cost value (absence of)	**<.001**	.029	1.000	1.000	**<.001**	1.000	**<.001**	**<.001**

Item/factor	Overall group difference		Paired group differences, sig. (p)					
	Sig. (p)	Size (η^2)	A-B	A-C	A-D	B-C	B-D	C-D
Orientation: mastery	**<.001**	.242	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Orientation: performance	**<.001**	.178	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Perceived control	**<.001**	.230	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Perceived control (exams)	**<.001**	.122	**<.001**	**<.001**	**.001**	**<.001**	**<.001**	**<.001**
Study strategy: self-regulation	**<.001**	.327	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Study strategy: control	**<.001**	.267	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Study strategy: memorisation	**<.001**	.229	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Study strategy: elaboration	**<.001**	.274	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Anxiety (absence of)	**<.001**	.154	**<.001**	**<.001**	**.001**	**<.001**	**<.001**	**<.001**
Mastery norms (good grade)	**.002**	.012	1.000	**.004**	1.000	.086	.866	**.007**
Subject-comparisons	**<.001**	.166	**<.001**	**<.001**	**<.001**	**.001**	**<.001**	**<.001**
Peer-comparisons	**<.001**	.116	**<.001**	**<.001**	**.001**	**<.001**	**<.001**	**<.001**
Social persuasions (praise)	**<.001**	.295	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Vicarious experiences	**<.001**	.141	**<.001**	**<.001**	**<.001**	**.023**	**<.001**	**<.001**
Norms/influence (friends)	**<.001**	.068	**<.001**	**<.001**	.072	**.013**	**<.001**	**<.001**
Norms/influence (parents)	**<.001**	.428	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Teacher perceptions	**<.001**	.272	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Teacher/school careers/events	**<.001**	.147	**<.001**	**<.001**	**<.001**	**.002**	**<.001**	**<.001**

Notes: Missing responses were estimated by expectation-maximisation. Significant p-values (p < .05) are highlighted in bold. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey; the clusters were formed on these factors.*

Section 8.3: Discussion

The analysis considered multiple complementary perspectives in order to address the second research question: whether students with different degrees of confidence accuracy/bias exhibited different science intentions, attitudes, and beliefs.

Initially, differences in intentions and attitudes were considered across conceptual accuracy/bias groups, formed through traditional methods, which are discussed and contextualised in **Section 8.3.1**. Differences were then considered across accuracy/bias clusters from latent-profile analysis, discussed in **Section 8.3.2**. As an alternate perspective, differences in accuracy/bias were considered across intention/attitude clusters from latent-profile analysis, discussed in **Section 8.3.3**. Conclusions for the second research question are then made in **Section 8.3.4**.

Fundamentally, the process offered the potential for disconfirmation and enhanced rigour: considering the accuracy/bias clusters helped consider whether the conceptual accuracy/bias groups were meaningful; considering the intention/attitude clusters helped consider whether accuracy/bias actually differed within meaningful contexts; and considering self-reflective accuracy/bias in the 2014/2015 survey helped consider whether the PISA 2006 results offered a meaningful insight into accuracy/bias.

Section 8.3.1: Accuracy/bias groups

Students were grouped into conceptual categories of confidence accuracy/bias, considering self-concept beliefs in the PISA 2006 survey and self-reflective task-confidence in the 2014/2015 survey. These conceptual groups considered under-confident, accurate, and over-confident students, where accurate students were separated into those with accurately-low or accurately-high confidence. This broadly extended earlier research, which had infrequently separated those with accurately-high or accurately-low confidence (Dupeyrat, Escribe, Huet, & Régner, 2011; Gonida & Leondari, 2011; Gresham, Lane, MacMillan, Bocian, & Ward, 2000).

Results from PISA 2006 for the above-average/below-average groups highlighted that students' intentions and attitudes (specifically, the key predictors of students' intentions) descended from the accurately-high group, then the over-confident group, then the under-confident group, and then the accurately-low group. Results from PISA 2006 for the regression-residual groups alternately highlighted that intentions and attitudes descended from the over-confident group, then the accurately-high group, then the accurately-low group, and then the under-confident group. Results from the 2014/2015 survey for the difference-score accuracy/bias groups highlighted that intentions and self-efficacy descended from the accurately-high group, then the over-confident and under-confident groups (which did not differ), and then the accurately-low group; for the other key predictors (interest value, utility value, personal value, and norms/influences from parents), attitudes descended from the accurately-high group, then the over-confident group, and then the under-confident and the accurately-low groups (which did not differ).

In general terms, these results somewhat supported the earlier hypothesis (**Section 4.2**) that over-confidence would associate with higher attitudes than under-confidence, but gave contrasting results regarding the potential benefits of over-confidence compared to accurately-high confidence. Prior research using regression-residual accuracy/bias groups has similarly highlighted that over-confident students reported higher interest than under-confident students, although also found that accurate students reported higher or similar interest to over-confident students, depending on the academic subject (Gonida & Leondari, 2011). The results from the 2014/2015 survey, directly considering task-level self-reflective accuracy/bias via a difference-score, highlighted that the under-confident and over-confident students reported similar orientations towards mastering learning and towards out-performing other students, while accurately-high students reported the highest orientations. This contrasted with the relevant hypothesis in **Section 4.2** and hence with earlier studies that found over-confident students to report higher out-performing orientations in some situations for secondary school students (Dupeyrat, Escribe, Huet, & Régner, 2011; Gonida & Leondari, 2011) and for university students (Bipp, Steinmayr, & Spinath, 2012; Willard & Gramzow, 2009). Future research may then need to explore task-level accuracy/bias in more detail, and/or consider any potential methodological impacts.

In the 2014/2015 survey, students with accurately-high task-level confidence reported higher current science grades, which cohered with earlier studies that have broadly associated higher accuracy with higher attainment (Chen, 2003; Rinne & Mazzocco, 2014). These students also reported higher self-regulated studying, which again broadly linked with established associations between attainment and self-regulated studying (Credé & Phillips, 2011; Sperling, Richmond, Ramsay, & Klapp, 2012). As a conceptual process, self-regulation may require accurate beliefs and/or generally entail that beliefs become accurate (Boekaerts, 1999; Butler & Winne, 1995; Pintrich, 1999); the idea of self-regulated studying, as involving cyclical phases of forethought, performance control, and self-reflection, may similarly be facilitated by accurate beliefs (Butler & Winne, 1995; Zimmerman, 2000; Zimmerman & Moylan, 2009). However, the idea of 'self-regulated studying' somewhat differs from the ideas of 'self-regulation' and confidence accuracy/bias, and further research would be required to clearly establish the relationship between them. The results also could not, in themselves, clarify whether reported self-regulated studying followed from accurate confidence, high confidence, and/or from high attainment.

Ultimately, the different approaches to forming groups entailed different conceptual definitions of accuracy/bias, but were each suited to their particular survey contexts. Specifically, PISA 2006 could only indirectly or implicitly consider accuracy/bias, hence regression-residual indicators were unavoidable; the simple cross-tabulation of above-average/below-average self-concept and task-scores also provided a plausible baseline comparison, and broadly matched the difference-score groups in the 2014/2015 survey. The 2014/2015 survey directly considered task-level accuracy/bias so a difference-score was more meaningful than a regression-residual indicator. The results fundamentally highlighted the need for attention (and clarity) regarding how accuracy/bias groups are defined, since different approaches can entail different interpretations regarding the relative benefits or detriments of the various groups.

Section 8.3.2: Accuracy/bias clusters

Considering clusters of students, revealed through latent-profile analysis, helped to determine the extents of under-confidence, accuracy, and over-confidence across the students in PISA 2006 and the 2014/2015 survey. The clusters were emergent from the data rather than relying on conceptual classifications, and could conceivably emerge with any size and/or with any degree of accuracy/bias.

There were some broad similarities between the clusters and the conceptual groups, for both surveys, suggesting that the clusters could be given the same descriptive labels as the groups (under-confident, accurately-low, accurately-high, and over-confident). The conceptual groups may therefore reflect naturally-emerging tendencies, but the meaning of the broad labels could differ when assigned to the groups or to the clusters; for example, the particular profile of an 'under-confident' cluster (e.g. average task-scores but slightly below-average self-concept beliefs in PISA 2006) differed from the conceptual 'under-confident' groups (e.g. above-average task-scores but below-average self-concept beliefs, or average task-scores but moderately/highly below-average self-concept beliefs). However, 'accurately-high' groups/clusters could also or conversely appear over-confident, hence results could vary in interpretation depending on the considered indicators (and the four broad labels may not necessarily then be ideal terms).

In the PISA 2006 survey, students' science intentions and attitudes were highest in the cluster with 'accurately-high' self-concept beliefs (cluster D, 10.4% of the considered students, with highly above-average self-concept beliefs and task-scores, but with relatively higher self-concept beliefs and hence some degree of over-confidence), then 'over-confident' students (C, 15.9%, students with moderately above-average task-scores and self-concept beliefs but with relatively higher self-concept beliefs), then 'under-confident' students (A, 55.9%, with average task-scores but slightly below-average self-concept beliefs), then 'accurately-low' students (B, 17.8%, with below-average task-scores and self-concept beliefs). In the 2014/2015 survey, students' intentions and attitudes were highest in the cluster with 'accurately-high' task-confidence (cluster C, 8.7% of the considered students), then ‘under-confident’ students (A, 42.3%, with moderately above-average task-scores and task-confidence but with relatively lower task-confidence), then ‘over-confident’ students (D, 7.5%, with below-average task-scores and task-confidence but with relatively higher task-confidence) and ‘accurately-low’ students (B, 41.5%) who generally reported similarly.

The PISA 2006 survey and the 2014/2015 survey therefore provided contrasting results concerning differences between the under-confident and over-confident clusters: PISA 2006 showed the same pattern across the conceptual groups and the emergent clusters, where over-confidence entailed higher attitudes than under-confidence; the results from the 2014/2015 survey suggested that under-confidence entailed higher attitudes than over-confidence when considering the clusters, but revealed little to no difference between the two biases when considering the conceptual groups. The groups and clusters, and the two surveys themselves, nevertheless involve different approaches (e.g. subject/task levels, relative/absolute differences), so some degree of difference is likely to be unavoidable. Additionally, interpretations depend on whether the groups/clusters with the highest attitudes are considered as accurate or over-confident; these groups/clusters could strengthen the apparent benefit of (some degree of relative) over-confidence.

Relatively little research has been undertaken in the area, so it remains difficult to contextualise these results. One example of research with secondary school students in Spain had revealed clusters of relatively equal sizes that were equivalent in profile to under-confident, accurately-low, accurately-high, and over-confident conceptual groups, essentially being akin to cross-tabulations of above-average/below-average mathematics performance and self-concept beliefs, for example where the cluster of under-confident students did indeed have above-average performance but below-average self-concept beliefs (Sáinz & Upadyaya, 2016). The PISA 2006 and 2014/2015 cluster results were therefore similar in potentially highlighting the same four tendencies, but differed in their cluster sizes and particular profiles. On a wider level, the PISA 2006 and 2014/2015 cluster results were also similar to prior findings where students with high confidence beliefs and high attitudes, identified through cluster analysis, reported studying more science courses (Simpkins & Davis-Kean, 2005), assuming that the students' science intentions would reflect their actual choices.

In both surveys, the cluster profiles suggested that students in higher-performing clusters had confidence (task-confidence or self-concept beliefs) in excess of their relative attainment (task-scores), although the pattern for the lower-performing clusters differed across the two surveys. PISA 2006 highlighted that the lowest-performing cluster was under-confident (via a regression-residual indicator), while the 2014/2015 survey highlighted that the lowest-performing cluster was over-confident (via a difference-score considering task-level self-reflective accuracy/bias). Considered in general terms, this partially reflected patterns seen in prior research with comparable student ages. Specifically, for secondary school students in Switzerland and Germany (Grade 9, age 15), for physics, the cluster with the highest performance had self-concept beliefs slightly higher than their performance, and the cluster with the lowest performance had self-concept beliefs slightly lower than their performance, when considered on equalised scales (Seidel, 2006). The results from the 2014/2015 survey clusters partly reflected those seen in studies with undergraduate students, which have used various methods and generally found that students with lower attainment exhibited over-confidence; however, undergraduate students with higher attainment have often exhibited a small degree of under-confidence, which contrasted with the results from PISA 2006 and the 2014/2015 survey (Ackerman & Wolman, 2007; Kruger & Dunning, 1999). The presented results from PISA 2006 and the 2014/2015 survey are nevertheless beneficial in providing initial perspectives for England, from which further research can then extend on.

Given the results from the accuracy/bias clusters, it may be somewhat difficult to conclusively determine whether under-confidence or over-confidence is detrimental or beneficial, since the clusters did not appear to exhibit extreme biases: some of the clusters could indeed be described as 'under-confident' and 'over-confident', but the magnitudes of these confidence biases appeared to be lower than those embodied by the (artificial) conceptual accuracy/bias groups.

Section 8.3.3: Intention/attitude clusters

The first approach to latent-profile analysis formed clusters on students' accuracy/bias via considering task-scores and self-concept beliefs or task-confidence, and then considered differences in the cluster profiles and proportions. The second approach to latent-profile analysis formed clusters on students' intentions and key predicting factors, and then considered differences in the cluster profiles, including their degrees of accuracy/bias.

Considered broadly, the fundamental findings cohered across both approaches. When considering clusters of students with different magnitudes of science intentions and attitudes, smaller clusters of students with the highest intentions and attitudes were revealed in the PISA 2006 and the 2014/2015 surveys; these clusters also exhibited the highest indicators of attainment (task-scores and/or science grades) and the highest indicators of accuracy in their beliefs. The students' confidence accuracy/bias indeed differed across the various intention/attitude clusters, highlighting that differences in accuracy/bias can be observed in meaningful contexts.

In PISA 2006, the intention/attitude clusters with below-average attitudes and beliefs exhibited relative under-confidence (via the regression-residual indicator) while clusters with above-average beliefs exhibited relative over-confidence. Relative 'over-confidence' perhaps depended on the particular regression-residual indicator, and may not necessarily entail biases in self-reflection. In the 2014/2015 survey, the cluster with the highest, above-average, intentions and attitudes also exhibited the highest self-reflective task-level accuracy and lowest degree of bias (i.e. under-confidence or over-confidence), although the (difference-score) accuracy/bias indicator suggested a slight degree of over-confidence. As before, these results broadly cohered with earlier studies that have associated higher accuracy with higher attainment (Chen, 2003; Rinne & Mazzocco, 2014), but contrasted with studies with undergraduate students that associated higher attainment with slight under-confidence and lower attainment with larger degrees of over-confidence (Ackerman & Wolman, 2007; Kruger & Dunning, 1999).

The intention/attitude clusters (and the earlier accuracy/bias clusters) appeared to exhibit broadly proportionate attitudes and beliefs within each cluster, which cohered with earlier studies (Chow, Eccles, & Salmela-Aro, 2012; Simpkins, Davis-Kean, & Eccles, 2006), although some within-cluster differences could perhaps still be seen. For example, in PISA 2006, the intention/attitude cluster with the lowest, below-average attitudes (cluster C in the four-cluster model) exhibited even lower science intentions than their other attitudes (which were more similar in magnitude); the cluster with the highest, above-average attitudes (cluster D) exhibited relatively higher perceived utility of science (and to a lesser degree, science intentions) than their other attitudes (which appeared more similar in magnitude). These patterns, however, did not clearly follow those seen in some prior studies, which have found, for example, distinct clusters of students exhibiting highly above-average confidence but moderately above-average utility and interest (Andersen & Chen, 2016), or exhibiting moderate self-concept beliefs and high interest (Aschbacher, Ing, & Tsai, 2014), together with other clusters. It remains relatively unclear to what extent results may vary across different samples, time-periods, and the considered factors; PISA reflected the situation as of 2006, for example, and educational contexts, and their yearly cohorts of students, change over time.

Within both surveys, relatively small proportions of students formed the clusters with the highest intentions and attitudes. Considering the four-cluster solutions, these covered 11.7% of students in PISA 2006 and 16.2% of students in the 2014/2015 survey; in the 2014/2015 survey, for example, such students had strongly above-average attitudes equivalent to ‘agree’ to ‘strongly agree’ on the underlying scale. Prior studies have similarly found that clusters of students with high confidence and attitudes (and/or intentions) have generally been small (Aschbacher, Ing, & Tsai, 2014; Institution of Mechanical Engineers, 2014). In broad terms, this supported the continuing assumptions and concerns around low numbers of students aspiring to study science subjects (Royal Society, 2008b). Additionally, the identification of distinct clusters highlighted that students’ attitudes and beliefs may be more closely associated than previously considered. It is possible that science identities, for example, may entail relatively high and cohering attitudes and beliefs.

Nevertheless, this does not mean that science education should only focus on those small numbers of students with universally high attitudes, which might only sustain patterns of under-representation (Claussen & Osborne, 2013; Osborne & Dillon, 2008). Instead, educators and researchers may need to consider how particular patterns of attitudes may form, and how students' attainment, confidence, and attitudes influence one another. Various interventions have indeed focused on students' attitudes towards science (Rosenzweig & Wigfield, 2016), including various approaches that have helped to increase students' interest in science (Bernacki, Nokes-Malach, Richey, & Belenky, 2016; Hong & Lin-Siegler, 2012; Hulleman & Harackiewicz, 2009). Similarly, promoting the utility of science for students and parents has associated with higher science interest and attainment for students, and with students selecting courses in science (Harackiewicz, Rozek, Hulleman, & Hyde, 2012; Rozek, Hyde, Svoboda, Hulleman, & Harackiewicz, 2015). Greater understanding of how students' attitudes associate may then inform whether interventions could feasibly continue to consider specific factors or intervention points, such as interest or utility alone, or whether interventions might need to consider many factors together or apply different approaches.

Section 8.3.4: Conclusions

Research question two: did students with different degrees of confidence accuracy/bias exhibit different science intentions, attitudes, and beliefs?

Fundamentally, the various accuracy/bias groups and accuracy/bias clusters did indeed exhibit different science attitudes and intentions. On a wider level, there were some similarities between the emergent clusters and the conceptual groups, suggesting that conceptual groups may indeed reflect natural tendencies towards accuracy and/or confidence biases. Similarly, students clustered on their intentions and attitudes (i.e. their intentions and key predicting factors of their intentions) exhibited different degrees of confidence accuracy/bias, which again provides reassurance that confidence accuracy/bias is contextually meaningful within science education.

Nevertheless, given that conceptual groups could entail different results depending on the particular grouping approach (i.e. the particular conceptual definitions of under-confidence and over-confidence that were applied), considering accuracy/bias clusters may be beneficial within future research; clusters can also help clarify the magnitudes and extents of particular tendencies towards accuracy/bias. Otherwise, it may be necessary to clarify the impact of different methodological approaches, and/or apply multiple approaches within one study.

In both surveys, students with accurately-high confidence generally reported the highest intentions and attitudes, across the conceptual groups and the emergent clusters, excepting that when considering the regression-residual conceptual accuracy/bias groups in PISA 2006 over-confident students reported the highest. Groups/clusters with 'accurately-high' confidence could also or conversely appear somewhat over-confident, depending on the considered indicators. This may suggest the potential 'benefits' of (relative) over-confidence, but perhaps highlight their contextual or conditional nature (perhaps requiring some level of underlying attainment and/or confidence).

Across the groups and clusters in PISA 2006, under-confident students exhibited lower attitudes than over-confident students, suggesting that under-confidence may be problematic; the students' attitudes and beliefs may closely associate, sufficient for their confidence to entail lower attitudes than might otherwise be expected, given their attainment. This was broadly supported by the accuracy/bias groups in the 2014/2015 survey: the under-confident and over-confident groups reported similar science intentions and self-efficacy, but under-confident students reported lower for other attitudes (including interest, utility, and personal value, found to be important predictors of science intentions) despite exhibiting higher attainment. However, the accuracy/bias clusters in the 2014/2015 survey conversely highlighted that those with over-confidence reported lower intentions and attitudes than those exhibiting under-confidence; the 'over-confident' cluster exhibited lower attainment than the over-confident conceptual group, however.

Despite such variation across approaches, given the limited prior research in England, the results provide a plausible and beneficial baseline for further exploration.

Section 9: Research question three

The third research question considered whether students with different degrees of confidence accuracy/bias considered their science intentions in different ways. The results from the PISA 2006 survey (**Section 9.1**) and the 2014/2015 survey (**Section 9.2**) are described in turn, and then discussed and contextualised in **Section 9.3**.

Section 9.1: Results: PISA 2006

Predictive modelling could not easily reveal the impact of accuracy/bias on students' intentions (**Section 7**); students were then categorised according to their confidence accuracy/bias, and their expressed intentions, attitudes, and other beliefs differed across the various groups and clusters (**Section 8**). The students' varying intentions may have followed from varying attitudes, and/or from students' applying varying processes of decision-making. Predictive modelling was then applied for the separate accuracy/bias groups (**Section 9.1.1**) and accuracy/bias clusters (**Section 9.1.2**) to help explore why the students expressed different science intentions. Predictive modelling could not feasibly be applied for the intention/attitude clusters, however, given that these clusters were already formed on the students' intentions and key predicting factors: within each cluster there would be less variance per factor, and less (if any) association between the factors, due to the nature of clustering the students.

Section 9.1.1: Accuracy/bias groups

Conceptual groups (above-average/below-average intersections)

The students' intentions and attitudes varied across the four (above-average/below-average) conceptual accuracy/bias groups (**Section 8.1.1**): those with accurately-high beliefs (33.2% of the considered students) reported the highest intentions and attitudes, then over-confident students

(18.2%), then under-confident students (20.1%), and then those with accurately-low beliefs (28.5%). The groups were formed so that under-confident students had below-average self-concept but above-average task-scores, while over-confident students had above-average self-concept but below-average task-scores.

Students' science intentions were indeed predicted by different factors in different ways across these four conceptual groups (**Table 29** summarises the comprehensive set of predictors). For those with accurately-high self-concept beliefs, science intentions were most strongly predicted by their utility value, personal value, interest value, and their interest in various science areas/topics. For those with over-confident beliefs, the strongest positive predictors were utility value, interest value, career information, and personal value, while applied teaching had a negative predictive association with intentions. For those with under-confident beliefs, the strongest predictors were utility value, interest value, and personal value. For those with accurately-low self-concept beliefs, the strongest positive predictors were utility value, personal value, careers information, self-concept beliefs, and interest value, while general value of science had a negative predictive association.

Numerous differences in coefficient magnitude occurred across the groups. The clearest difference was that the students' perceived utility of science was predictive of science intentions for all groups, but was relatively stronger for those students with accurately-high self-concept beliefs. Various other differences were apparent. For example, interest value had a lower predictive magnitude for accurately-low students compared to under-confident students. Careers information had a higher predictive magnitude for over-confident students than accurately-high and under-confident students; careers information also had a higher predictive magnitude for accurately-low students than under-confident students.

As examples of sensitivity checks, preliminary analysis also involved predicting science intentions using only the key predictors highlighted across the entire sample (**Section 7.1**), specifically using students' perceived utility of science, interest/enjoyment of science, personal value of science, and their science self-concept beliefs, but also including task-scores and students' background. When considering these smaller sets of predictors, the majority of the same results were seen excepting that no coefficient differences were observed across the accurately-high and the under-confident groups (which were both formed with above-average task-scores but with either above-average or below-average self-concept beliefs). In some cases, therefore, it may still be difficult to isolate effects following from the magnitude of task-scores and/or self-concept from effects following from the degree of accuracy/bias. As highlighted across the entire sample (**Section 7.1**), various factors may mediate the associations between other predictors and science intentions, which may introduce slightly varying results depending on which factors are considered, which may complicate any modelling.

Fundamentally, when modelling the comprehensive array of predictors (**Table 29**), there were coefficient differences across all possible pairs of groups, although not necessarily for each predictor. There were still sufficient differences in predictive coefficients across these conceptual groups to infer that the students' accuracy and/or bias appeared to be relevant moderators of the predictive associations between various factors and students' science intentions, especially for utility value, interest value, career information, teaching via practical/hands-on activities, and other factors.

Table 29: PISA (England) 2006: conceptual groups (above-average/below-average self-concept/task-score intersections), predicting science intentions

	Under-confident (U)		Accurately-low (L)		Accurately-high (H)		Over-confident (O)	
Item/factor	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)
Intercept/constant	[UO] **NA**	**.029**	**NA**	**.023**	[HO] **NA**	**<.001**	[UO HO] NA	.564
Gender (1=boy)	[UL] -.016	.527	[UL LO] **.054**	**.021**	[HO] .020	.216	[LO HO] **-.070**	**.011**
Books at home	-.038	.157	-.017	.504	-.007	.695	-.013	.652
Parent(s) in science (1=yes)	[UH] .036	.149	-.013	.588	[UH] -.020	.207	.021	.444
Mothers' education	-.037	.184	.021	.443	.004	.841	.012	.705
Fathers' education	-.001	.958	-.011	.684	.013	.488	.017	.594
Task-score (PV1)	[UO] **.058**	**.030**	[LH] .007	.785	[LH HO] **.070**	**<.001**	[UO HO] -.045	.149
Self-concept	.049	.082	[LH] **.122**	**<.001**	[LH] .036	.058	.040	.176
Self-efficacy (areas)	-.027	.337	**-.072**	**.011**	**-.042**	**.020**	-.044	.145
Interest (various areas)	**.069**	**.032**	[LH] **.083**	**.005**	[LH HO] **.112**	**<.001**	[HO] **.070**	**.039**
Interest value	[UL] **.197**	**<.001**	[UL] **.108**	**<.001**	**.154**	**<.001**	**.187**	**<.001**
Utility value	[UL UH UO] **.427**	**<.001**	[UL LH] **.295**	**<.001**	[UH LH HO] **.480**	**<.001**	[UO HO] **.318**	**<.001**
Personal value	**.134**	**<.001**	**.154**	**<.001**	**.212**	**<.001**	**.149**	**<.001**
General value	**-.073**	**.009**	**-.115**	**<.001**	**-.092**	**<.001**	-.035	.358
Science activities	**.056**	**.042**	**.087**	**.001**	**.049**	**.013**	.044	.202
School career preparation	.001	.960	-.030	.268	-.027	.148	**-.073**	**.022**
School career information	[UL UO] .014	.596	[UL] **.130**	**<.001**	[HO] **.063**	**<.001**	[UO HO] **.181**	**<.001**
Teaching: interaction	-.016	.593	**-.069**	**.028**	-.036	.072	.008	.847
Teaching: activities	[UH] -.022	.431	[LH] .014	.642	[UH LH HO] **-.097**	**<.001**	[HO] -.020	.612
Teaching: investigations	<.001	.987	**.058**	**.043**	<.001	.988	.029	.444
Teaching: applications	-.027	.391	-.048	.141	-.039	.062	**-.102**	**.018**
Explained variance	48.6%		36.4%		64.6%		42.9%	
Unexplained variance (residual)	50.9%		62.4%		34.6%		54.8%	
Unexplained variance (school)	.5%		1.2%		.8%		2.3%	

Notes: Missing responses were estimated by expectation-maximisation. Standardised coefficients ('Std. Effect') and significance (p-values; 'Sig. (p)') are reported. Significant coefficients (p < .05) are highlighted in bold. Significant differences in coefficient magnitude (p < .05) across groups are highlighted in superscript.

Conceptual groups (regression-residual accuracy/bias groups)

The students' intentions and attitudes varied across the four (regression-residual) conceptual accuracy/bias groups (**Section 8.1.1**): those with over-confident beliefs (26.8% of the considered students) reported the highest intentions and attitudes, then those with accurately-high beliefs (23.6%), then those with accurately-low beliefs (20.6%), and then those with under-confident beliefs (28.9%). The groups were formed via classifying the regression-residual indicator of accuracy/bias, which essentially ensured that under-confident students had below-average self-concept and average task-scores, while over-confident students had above-average self-concept and average task-scores.

Students' science intentions were predicted by different factors in different ways across these regression-residual groups (**Table 30** summarises the comprehensive set of predictors). For example, utility value had a relatively higher predictive association with science intentions for accurately-high students than for the other groups. Compared to accurately-high students, interest in areas/topics within science had a relatively lower predictive association for under-confident students. The students' personal value of science and reported careers information had lower predictive associations with science intentions for under-confident students than for over-confident students; for under-confident students, the reported careers information had a sufficiently low magnitude as to be insignificantly predictive (i.e. essentially not differing from zero).

Students' task-scores were predictive of their intentions only for under-confident students, while students' self-concept beliefs were predictive only for accurately-low students, although the coefficient magnitudes were only significantly different when considering a smaller set of key predictors (from preliminary/sensitivity analysis and not tabulated for brevity) and not the comprehensive set of predictors (**Table 30**); accordingly, such results may not necessarily be conclusive. Nevertheless, the other patterns of coefficients (such as for utility and personal value) across the groups were not dependent on whether smaller or larger sets of predictors were used.

Fundamentally, the regression-residual groups (**Table 30**) and the earlier above-average/below-average groups (**Table 29**) showed some similarities, despite the different methods in forming the groups. For example, careers information had a significantly higher predictive association with science intentions for both of the over-confident groups, and for both of the accurately-low groups, when compared to their respective under-confident groups. Utility value was also more predictive of science intentions for accurately-high students than those in the other groups across both approaches.

However, differences across the methods/approaches were also seen. Considering the above-average/below-average groups (**Table 29**), utility value had a higher predictive association for under-confident students than over-confident students, while no differences in predictive associations across the groups were observed for students' personal value of science. Conversely, considering the regression-residual groups (**Table 30**), personal value had a lower predictive association with science intentions for under-confident students than for over-confident students, while no difference across these two groups was observed for utility value. Accordingly, while similarities were observed, different methods of forming accuracy/bias groups could potentially entail some different conclusions/inferences, such as whether to focus on either utility value or personal value when further exploring the implications of under-confidence and over-confidence within science education.

Table 30: PISA (England) 2006: conceptual groups (regression-residual accuracy/bias groups, ±.5 SD boundaries), predicting science intentions

	Under-confident (U)		Accurately-low (L)		Accurately-high (H)		Over-confident (O)	
Item/factor	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)
Intercept/constant	**NA**	**.001**	NA	.295	**NA**	**.009**	**NA**	**.004**
Gender (1=boy)	.014	.532	.028	.307	.013	.530	-.017	.396
Books at home	-.037	.135	.002	.935	.014	.540	-.010	.663
Parent(s) in science (1=yes)	.017	.434	-.028	.316	-.009	.671	.026	.197
Mothers' education	.003	.914	-.021	.532	-.020	.394	.028	.224
Fathers' education	-.030	.233	.038	.265	.024	.321	<.001	.990
Task-score (PV1)	**.081**	**.009**	-.006	.862	.053	.054	.043	.166
Self-concept	[UL] .053	.074	[UL LO] **.129**	**<.001**	.049	.074	[LO] .043	.083
Self-efficacy (areas)	[UL] -.016	.563	[UL LH LO] **-.140**	**<.001**	[LH] **-.046**	**.047**	[LO] -.030	.225
Interest (various areas)	[UH] .055	.052	[LH] **.073**	**.031**	[UH LH] **.121**	**<.001**	**.067**	**.009**
Interest value	**.167**	**<.001**	**.147**	**<.001**	**.166**	**<.001**	**.189**	**<.001**
Utility value	[UH] **.386**	**<.001**	[LH LO] **.337**	**<.001**	[UH LH HO] **.468**	**<.001**	[LO HO] **.373**	**<.001**
Personal value	[UO] **.122**	**<.001**	[LO] **.142**	**<.001**	**.172**	**<.001**	[UO LO] **.251**	**<.001**
General value	**-.104**	**<.001**	**-.071**	**.049**	**-.088**	**<.001**	**-.101**	**<.001**
Science activities	**.078**	**.002**	**.066**	**.039**	**.049**	**.039**	.046	.071
School career preparation	-.006	.823	**-.071**	**.018**	-.037	.097	-.029	.240
School career information	[UL UO] .041	.091	[UL] **.157**	**<.001**	**.075**	**.001**	[UO] **.108**	**<.001**
Teaching: interaction	-.021	.443	-.061	.091	**-.055**	**.023**	-.026	.336
Teaching: activities	[UH] -.019	.476	[LH LO] .052	.140	[UH LH] **-.094**	**<.001**	[LO] **-.076**	**.004**
Teaching: investigations	.001	.973	.059	.083	.010	.669	.017	.498
Teaching: applications	[UL] -.011	.698	[UL] **-.100**	**.008**	-.034	.169	-.046	.111
Explained variance	42.5%		33.9%		59.0%		57.0%	
Unexplained variance (residual)	55.8%		65.8%		41.0%		41.8%	
Unexplained variance (school)	1.7%		.3%		.0%		1.2%	

Notes: Missing responses were estimated by expectation-maximisation. Standardised coefficients ('Std. Effect') and significance (p-values; 'Sig. (p)') are reported. Significant coefficients (p < .05) are highlighted in bold. Significant differences in coefficient magnitude (p < .05) across groups are highlighted in superscript.

Section 9.1.2: Accuracy/bias clusters

The students' reported science intentions and attitudes differed across the four accuracy/bias clusters from latent-profile analysis (**Section 8.1.2**). Essentially, intentions and attitudes were highest in the cluster with 'accurately-high' self-concept beliefs (cluster D, 10.4% of the considered students, with highly above-average self-concept beliefs and task-scores, but with some degree of over-confidence), then 'over-confident' students (cluster C, 15.9%, students with moderately above-average task-scores and self-concept beliefs but with relatively higher self-concept beliefs), then 'under-confident' students (cluster A, 55.9%, with average task-scores but slightly below-average self-concept beliefs), then 'accurately-low' students (cluster B, 17.8%, with below-average task-scores and self-concept beliefs).

Students' science intentions were predicted in different ways across these different clusters (**Table 31**). Some similarities were also observed in the patterns of coefficients across these accuracy/bias clusters and the earlier accuracy/bias groups (**Section 9.1.1**). While the students' perceived utility of science and personal value of science were predictive of science intentions for all clusters, the predictive associations were lowest for those students in cluster B, especially compared to students in cluster D; this partially mirrored the differences between the 'accurately-low' and 'accurately-high' groups (although recognising that the personal value coefficients did not statistically-significantly differ for the groups). The personal value of science had a relatively lower predictive coefficient for cluster A compared to cluster C, which mirrored the difference between the 'under-confident' and 'over-confident' regression-residual groups. Reporting more student-led investigations within science lessons/teaching associated with higher science intentions only for students in cluster B (the factor was not significantly predictive for the other clusters), which was also seen for the 'accurately-low' above-average/below-average group. Students in clusters B and D reported, on average, experiencing similar levels of student-led investigations (**Table 16** and **Table 17**), which perhaps emphasises the importance of considering whether different students might consider different factors in different ways during their decision-making. Various other differences were also apparent; for example, students' task-scores were more predictive of science intentions for those students in cluster D than for those in other clusters. However, many differences across the clusters did not clearly correspond to those observed across the accuracy/bias groups, which perhaps unavoidably reflected the differences in profiles (and student numbers) across the groups and the clusters.

Table 31: PISA (England) 2006: accuracy/bias clusters (four clusters), predicting science intentions

	Cluster A		Cluster B		Cluster C		Cluster D	
Item/factor	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)
Intercept/constant	[AB] **NA**	**<.001**	[AB BD] NA	.751	[CD] NA	.077	[BD CD] **NA**	**<.001**
Gender (1=boy)	.009	.552	-.004	.894	.007	.795	-.020	.507
Books at home	[AD] -.001	.947	-.026	.345	[CD] <.001	.995	[AD CD] **-.090**	**.005**
Parent(s) in science (1=yes)	-.013	.404	.043	.091	.037	.168	-.027	.371
Mothers' education	[AC] -.017	.333	.010	.743	[AC] .057	.062	.004	.891
Fathers' education	.018	.307	.006	.840	-.032	.308	.018	.591
Task-score (PV1)	[AB AD] **.060**	**.001**	[AB BC BD] -.052	.090	[BC CD] .065	.066	[AD BD CD] **.140**	**<.001**
Self-concept	**.053**	**.001**	**.201**	**<.001**	-.048	.060	.015	.618
Self-efficacy (areas)	[AB] **-.076**	**<.001**	[AB] .006	.858	-.045	.147	-.042	.177
Interest (various areas)	[AB] **.091**	**<.001**	[AB BD] .037	.302	**.073**	**.032**	[BD] **.128**	**<.001**
Interest value	[AB] **.178**	**<.001**	[AB] **.134**	**<.001**	**.143**	**<.001**	**.121**	**.003**
Utility value	[AB] **.409**	**<.001**	[AB BC BD] **.229**	**<.001**	[BC] **.419**	**<.001**	[BD] **.440**	**<.001**
Personal value	[AC] **.155**	**<.001**	[BC BD] **.144**	**.001**	[AC BC] **.241**	**<.001**	[BD] **.249**	**<.001**
General value	**-.087**	**<.001**	**-.080**	**.031**	**-.106**	**.002**	**-.085**	**.019**
Science activities	**.062**	**<.001**	**.082**	**.012**	.048	.139	.062	.104
School career preparation	**-.037**	**.030**	-.006	.859	-.045	.138	.005	.892
School career information	**.089**	**<.001**	**.129**	**<.001**	**.084**	**.005**	.060	.060
Teaching: interaction	-.027	.143	-.051	.170	**-.074**	**.040**	-.030	.453
Teaching: activities	**-.049**	**.006**	-.012	.742	-.041	.239	-.037	.308
Teaching: investigations	[AB AC] .018	.301	[AB BC BD] **.104**	**.004**	[AC BC] -.062	.074	[BD] -.024	.494
Teaching: applications	-.035	.076	-.054	.168	.002	.955	**-.086**	**.028**
Explained variance	45.2%		52.3%		53.9%		62.6%	
Unexplained variance (residual)	54.4%		47.0%		45.1%		34.2%	
Unexplained variance (school)	.4%		.6%		1.0%		3.2%	

Notes: Missing responses were estimated by expectation-maximisation. Standardised coefficients ('Std. Effect') and significance (p-values; 'Sig. (p)') are reported. Significant coefficients (p < .05) are highlighted in bold. Significant differences in coefficient magnitude (p < .05) across clusters are highlighted in superscript.

Section 9.2: Results: 2014/2015 survey

Predictive modelling was applied for each of the various groups and clusters considered in the 2014/2015 survey, as with the PISA 2006 survey. This involved considering the accuracy/bias groups (**Section 9.2.1**) and the accuracy/bias clusters (**Section 9.2.2**) to help explore why the students expressed different science intentions.

Section 9.2.1: Accuracy/bias groups

Conceptual groups (difference-score groups with accurately-high and accurately-low groups)

Examining the conceptual (difference-score) accuracy/bias groups in the 2014/2015 survey, where the students' task-level self-reflective accuracy/bias was measured, highlighted that (**Section 8.2.1**): the students' intentions and self-efficacy descended from the accurately-high group, then the over-confident and under-confident groups (which did not differ), and then the accurately-low group; alternately, the other key predictors of intentions (interest value, utility value, personal value, and norms/influences from parents) descended from the accurately-high group, then the over-confident group, and then the under-confident and the accurately-low groups (which did not differ). The groups were relatively similarly sized, with 29.4% of the considered students across Years 9, 10, and 11 classified as under-confident, 21.2% as accurately-low, 21.4% as accurately-high, and 28.1% as over-confident.

Fewer students were considered in the 2014/2015 survey compared to PISA 2006, so that undertaking predictive modelling using many predictors (estimating many parameters) for each group/cluster could potentially increase the uncertainty associated with the parameter estimates, given smaller groups/clusters. Accordingly, preliminary/sensitivity analysis considered smaller and larger set of predictors, covering theorised factors and those previously highlighted as key predictors (**Section 7.2**); the fundamental results were nevertheless broadly similar across the various models.

When the students' science intentions were predicted for each group, various predictive coefficients significantly differed across the groups (**Table 32** shows the comprehensive set of predictors and **Table 33** shows a smaller set of predictors for illustration). For example, science intentions were more strongly predicted by perceived utility of science, self-efficacy, and the students' academic year itself, for accurately-high students compared to under-confident students. Conversely, science intentions were more strongly predicted by personal value of science for under-confident students than for accurately-high students; surprisingly, the students' personal value of science was not significantly predictive for the accurately-high students.

Research and commentary within science education has often assumed that students' aspirations and attitudes to science decline as they grow older (Osborne, Simon, & Collins, 2003; Royal Society, 2008b). The students' reported attitudes and beliefs, on average, appeared to support such an assumption, although this could not be conclusively determined due to the low numbers involved and since the students were not surveyed over time (the means per academic year are detailed in **Appendix 7** for reference). Considered across all students in Years 9, 10, and 11, the indicator of academic year was not predictive of science intentions, once the students' attitudes were included as predictors (**Table 8**). However, when considering the accuracy/bias groups separately, the indicator of the students' academic year was positively predictive of science intentions for accurately-high students, when controlling for their various attitudes and backgrounds (**Table 32**). Further (unknown) factors would need to be considered in order to help explain such a result, and/or exploration undertaken into whether, for some specific groups of students at least, science intentions actually increase over time (which is perhaps plausible, given prior research as reviewed in **Section 2.2.1**).

Table 32: England 2014/2015 (Years 9, 10, 11): difference-score accuracy/bias groups with accurately-low/accurately-high groups, predicting science intentions (comprehensive set of predictors)

	Under-confident (U)		Accurately-low (L)		Accurately-high (H)		Over-confident (O)	
Item/factor	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)
Intercept	[UH] NA	.513	[LH] NA	.216	[UH LH HO] **NA**	**<.001**	[HO] NA	.323
Year	[UH] -.017	.610	[LH] -.073	.183	[UH LH HO] **.147**	**.002**	[HO] .025	.523
Gender (1=boy)	.005	.882	-.029	.509	-.022	.586	.024	.514
Ethnicity (Black)	[UL] -.017	.566	[UL] .093	.052	.019	.605	-.009	.799
Ethnicity (East-Asian)	.055	.076	.042	.330	-.021	.602	-.026	.452
Ethnicity (South-Asian/Indian)	.012	.744	**.104**	**.032**	.044	.410	.075	.059
Ethnicity (mixed)	-.005	.862	.038	.393	.034	.362	.068	.051
Ethnicity (other)	.030	.300	.065	.132	-.009	.810	.035	.285
Books at home	-.020	.550	-.087	.065	-.022	.588	-.042	.299
Parent(s) in science (1=yes)	.010	.746	.054	.247	-.022	.558	.047	.185
Mothers' education	-.044	.252	.004	.948	.042	.380	.041	.410
Fathers' education	.033	.429	-.055	.372	-.085	.108	-.057	.258
Task-score	.037	.448	[LH] -.129	.162	[LH] .105	.187	.021	.689
Task-confidence	.014	.776	-.005	.957	.020	.806	.049	.377
Current grade	-.080	.073	-.034	.537	-.020	.737	-.056	.287
Self-concept	[UH] **.107**	**.032**	.096	.127	[UH] -.058	.303	-.006	.917
Self-efficacy	[UH] .037	.421	**.140**	**.011**	[UH] **.176**	**.002**	.102	.069
Interest value	.068	.262	.067	.358	.022	.741	**.152**	**.023**
Utility value	[UL UH] **.531**	**<.001**	[UL LH LO] **.185**	**.019**	[UH LH HO] **.701**	**<.001**	[LO HO] **.375**	**<.001**
Personal value	[UH] **.249**	**<.001**	**.218**	**.001**	[UH] .045	.432	**.205**	**.001**
Cost value (absence of)	**.070**	**.029**	.056	.277	-.032	.419	.067	.125
Orientation: mastery	-.037	.315	.024	.650	-.065	.148	-.068	.188
Orientation: performance	-.026	.434	**-.098**	**.048**	-.079	.064	-.056	.242
Perceived control	[UH] **-.124**	**.002**	-.019	.744	[UH] .074	.081	-.040	.465
Perceived control (exams)	[UL UH] .035	.352	[UL] **-.106**	**.037**	[UH] -.090	.059	-.073	.091
Study strategy: self-regulation	.051	.353	-.072	.326	-.022	.711	-.040	.554

Item/factor	Under-confident (U) Std. Effect	Under-confident (U) Sig. (*p*)	Accurately-low (L) Std. Effect	Accurately-low (L) Sig. (*p*)	Accurately-high (H) Std. Effect	Accurately-high (H) Sig. (*p*)	Over-confident (O) Std. Effect	Over-confident (O) Sig. (*p*)
Study strategy: control	.045	.377	[LH] **.146**	**.045**	[LH] -.051	.381	.068	.272
Study strategy: memorisation	[UO] -.040	.440	.126	.070	.041	.398	[UO] **.163**	**.007**
Study strategy: elaboration	-.038	.368	.075	.254	[HO] .083	.094	[HO] -.077	.160
Anxiety (absence of)	.058	.218	.043	.537	.010	.844	.020	.705
Mastery norms (good grade)	.046	.145	-.007	.883	-.023	.609	-.025	.512
Subject-comparisons	.003	.937	.026	.677	.070	.142	**.119**	**.017**
Peer-comparisons	.020	.619	-.053	.382	-.011	.835	.016	.744
Social persuasions (praise)	**-.105**	**.009**	**-.134**	**.018**	.005	.920	-.019	.739
Vicarious experiences	[UH] -.009	.795	**-.106**	**.035**	[UH] **-.139**	**<.001**	**-.087**	**.048**
Norms/influence (friends)	-.010	.754	[LH] **-.119**	**.016**	[LH] .026	.509	**-.091**	**.032**
Norms/influence (parents)	.064	.120	**.154**	**.017**	**.123**	**.013**	**.187**	**<.001**
Teacher perceptions	-.060	.175	.034	.613	-.048	.379	-.038	.555
Teacher/school careers/events	[UH] **.075**	**.043**	[LH LO] **.144**	**.009**	[UH LH] -.034	.462	[LO] -.012	.815
Explained variance	68.4%		52.4%		65.7%		69.2%	
Unexplained variance (residual)	31.6%		47.0%		32.4%		30.8%	
Unexplained variance (school)	.0%		.6%		1.8%		.0%	

Notes: Missing responses were estimated by expectation-maximisation. Standardised coefficients ('Std. Effect') and significance (p-values; 'Sig. (p)') are reported. Significant coefficients (p < .05) are highlighted in bold. Significant differences in coefficient magnitude (p < .05) across groups are highlighted in superscript.

Table 33: England 2014/2015 (Years 9, 10, 11): difference-score accuracy/bias groups with accurately-low/accurately-high groups, predicting science intentions (smaller set of predictors)

	Under-confident (U)		Accurately-low (L)		Accurately-high (H)		Over-confident (O)	
Item/factor	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)
Intercept	[UH] NA	.695	[LH] NA	.249	[UH LH HO] **NA**	**<.001**	[HO] NA	.427
Year	[UH] -.025	.415	[LH] -.072	.173	[UH LH HO] **.135**	**.002**	[HO] .009	.799
Gender (1=boy)	.012	.710	-.049	.247	-.015	.684	.001	.967
Task-score	[UL] .058	.191	[UL LH] -.143	.097	[LH] .108	.150	.008	.883
Task-confidence	.001	.978	.031	.712	-.010	.891	.026	.625
Current grade	-.051	.184	.003	.949	-.047	.368	-.020	.677
Self-concept	.082	.064	.089	.134	-.024	.634	.065	.186
Self-efficacy	[UH] .049	.251	**.128**	**.011**	[UH] **.172**	**.001**	.096	.059
Interest value	.061	.251	.064	.326	.028	.638	.104	.093
Utility value	[UL UH UO] **.543**	**<.001**	[UL LH] **.226**	**.003**	[UH LH HO] **.703**	**<.001**	[UO HO] **.316**	**<.001**
Personal value	[UH] **.263**	**<.001**	[LH] **.244**	**<.001**	[UH LH HO] .059	.278	[HO] **.232**	**<.001**
Cost value (absence of)	[UH] **.072**	**.019**	.064	.204	[UH] -.032	.395	.044	.277
Orientation: mastery	-.030	.388	[LH] .046	.343	[LH] **-.091**	**.026**	-.088	.070
Orientation: performance	-.030	.345	-.066	.164	**-.085**	**.033**	-.016	.711
Perceived control	[UL UH] **-.135**	**<.001**	[UL] -.013	.801	[UH] .055	.159	-.060	.198
Perceived control (exams)	[UL UH UO] .027	.449	[UL] **-.124**	**.008**	[UH] **-.111**	**.009**	[UO] **-.099**	**.013**
Study strategy: memorisation	[UL UO] -.018	.603	[UL LH] **.184**	**<.001**	[LH] .041	.282	[UO] **.112**	**.014**
Anxiety (absence)	.066	.083	.016	.754	.042	.336	**.091**	**.026**
Social persuasions (praise)	**-.112**	**.003**	**-.137**	**.010**	.002	.968	-.033	.533
Vicarious experiences	[UL UH] .006	.851	[UL] **-.120**	**.011**	[UH] **-.130**	**<.001**	-.077	.066
Norms/influence (friends)	-.015	.609	**-.108**	**.022**	-.015	.668	-.086	.034
Norms/influence (parents)	[UO] .067	.077	.105	.081	**.114**	**.010**	[UO] **.221**	**<.001**
Career/events	[UL UH] .058	.068	[UL LH LO] **.181**	**<.001**	[UH LH] -.048	.219	[LO] .002	.956
Explained variance	69.2%		51.4%		66.5%		59.4%	
Unexplained variance (residual)	30.8%		47.4%		32.1%		40.6%	
Unexplained variance (school)	.0%		1.2%		1.5%		.0%	

Notes: Missing responses were estimated by expectation-maximisation. Standardised coefficients ('Std. Effect') and significance (p-values; 'Sig. (p)') are reported. Significant coefficients (p < .05) are highlighted in bold. Significant differences in coefficient magnitude (p < .05) across groups are highlighted in superscript.

Section 9.2.2: Accuracy/bias clusters

Using latent-profile analysis to reveal clusters of students in the 2014/2015 survey, considering their task-confidence and task-scores and considering students across Years 9, 10, and 11, highlighted that (**Section 8.2.2**): students' intentions and attitudes were highest in the cluster with 'accurately-high' task-confidence (cluster C, 8.7% of the considered students), then 'under-confident' students (cluster A, 42.3%, with moderately above-average task-scores and task-confidence but with relatively lower task-confidence), then 'over-confident' students (cluster D, 7.5%, with below-average task-scores and task-confidence but with relatively higher task-confidence) and 'accurately-low' students (cluster B, 41.5%) who generally reported similarly.

As with the accuracy/bias groups, when predicting the students' science intentions for each cluster, the lower numbers of students in some clusters may have reduced the reliability and/or entailed that considering higher numbers of predictors would be less feasible. Accordingly, preliminary/sensitivity analysis considered comprehensive and reduced sets of predictors; the patterns of differences across the clusters were similar for many predictors, but not all. It remains unclear whether inconsistencies when modelling different sets of predictors were meaningful (perhaps highlighting potential mediation between factors that would need to be explored further) or reflected the inherent difficulty of modelling clusters with smaller numbers of students; further research with increased numbers of students would be necessary to clarify the area.

Fundamentally, when the students' science intentions were predicted for each cluster, various predictive coefficients significantly differed across the clusters (**Table 34** shows the comprehensive set of predictors and **Table 35** shows a smaller set of predictors for illustration). For example, the predictive association between utility value and science intentions varied across some of the pairs of clusters, although the patterns varied depending on which predictors were included.

Some patterns of coefficients and differences broadly followed those seen when considering the accuracy/bias groups (**Section 9.2.1**). For example, for cluster B (and as previously seen in the 'accurate-low' group), students' science intentions were predicted by their perceived utility of science at a lower magnitude, and were predicted by the reported provision of careers from teachers or the school at a higher magnitude, when compared to some of the other clusters. For cluster C (and as previously seen in the 'accurately-high' group), the students' intentions were most strongly predicted by their perceived utility of science and by their self-efficacy, although perhaps due to the greater uncertainty due to the smaller number of students in some clusters, only the predictive magnitude of perceived utility differed across the clusters in **Table 35** when considering a reduced set of predictors. For those in cluster C, the indicator of the students' academic year was more strongly (and positively) predictive of the students' science intentions, and differed in predictive magnitude compared to other clusters, but only when considering the full set of predictors (**Table 34**). However, the pattern of coefficients for cluster D had few similarities with those seen for the conceptual 'over-confident' group, and cluster A had few similarities with the conceptual 'under-confident' group, which may have reflected the varying profiles and/or numbers of students being considered.

Table 34: England 2014/2015 (Years 9, 10, 11): task-level accuracy/bias clusters (four clusters), predicting science intentions (comprehensive set of predictors)

	Cluster A		Cluster B		Cluster C		Cluster D	
Item/factor	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)
Intercept	[AB AC AD] **NA**	**.007**	[AB BC] NA	.484	[AC BC CD] **NA**	**<.001**	[AD CD] NA	.075
Year	[AB] .052	.075	[AB] -.053	.147	.141	.173	-.097	.277
Gender (1=boy)	<.001	.995	.029	.364	-.054	.544	-.085	.241
Ethnicity (Black)	.007	.754	.040	.229	.002	.978	.076	.384
Ethnicity (East-Asian)	.028	.266	-.005	.883	.023	.815	.002	.971
Ethnicity (South-Asian/Indian)	.036	.276	**.085**	**.017**	.072	.568	-.012	.897
Ethnicity (mixed)	-.001	.971	.050	.107	.110	.157	.039	.635
Ethnicity (other)	.021	.382	.058	.052	.027	.724	-.002	.972
Books at home	-.031	.260	-.049	.139	.092	.235	-.049	.577
Parent(s) in science (1=yes)	-.006	.808	.044	.167	-.007	.920	.009	.902
Mothers' education	[AD] .021	.486	-.023	.578	[CD] .120	.243	[AD CD] **-.266**	**.022**
Fathers' education	.005	.890	-.036	.388	-.107	.313	.173	.191
Task-score	.034	.210	-.013	.705	.064	.413	-.102	.242
Task-confidence	.015	.625	-.033	.353	.003	.968	-.004	.964
Current grade	-.022	.537	-.046	.244	.064	.552	.043	.718
Self-concept	.057	.146	.042	.348	.048	.675	-.077	.538
Self-efficacy	**.085**	**.021**	**.089**	**.034**	**.189**	**.035**	**.223**	**.043**
Interest value	**.122**	**.011**	.073	.175	-.069	.626	**.296**	**.027**
Utility value	[AB AD] **.579**	**<.001**	[AB] **.399**	**<.001**	**.500**	**<.001**	[AD] .292	.096
Personal value	**.174**	**<.001**	**.161**	**.002**	.179	.102	.041	.790
Cost value (absence of)	.016	.566	**.097**	**.008**	-.069	.347	-.153	.198
Orientation: mastery	[AD] -.030	.301	[BD] -.019	.635	-.099	.337	[AD BD] **-.317**	**.005**
Orientation: performance	**-.056**	**.047**	-.056	.128	-.090	.315	.073	.480
Perceived control	-.001	.972	[BC] **-.098**	**.021**	[BC] **.201**	**.016**	.076	.542
Perceived control (exams)	**-.091**	**.005**	-.048	.182	-.004	.971	-.031	.740
Study strategy: self-regulation	.015	.734	.044	.424	-.060	.623	-.121	.344
Study strategy: control	-.037	.366	.076	.150	-.148	.238	.098	.488

	Cluster A		Cluster B		Cluster C		Cluster D	
Item/factor	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)
Study strategy: memorisation	**.073**	**.048**	.053	.300	.049	.598	.189	.276
Study strategy: elaboration	-.052	.136	<.001	.999	.151	.090	-.022	.878
Anxiety (absence of)	.017	.660	.033	.484	-.084	.355	.092	.471
Mastery norms (good grade)	[AC] -.027	.320	.020	.546	[AC CD] **.177**	**.038**	[CD] -.078	.416
Subject-comparisons	.046	.158	.060	.168	-.017	.827	.114	.255
Peer-comparisons	-.007	.826	-.036	.383	.132	.188	-.046	.742
Social persuasions (praise)	-.048	.146	-.081	.057	-.052	.586	-.015	.907
Vicarious experiences	[AC] -.018	.510	**-.106**	**.004**	[AC] **-.237**	**.004**	-.012	.911
Norms/influence (friends)	-.041	.115	**-.069**	**.049**	.122	.153	-.023	.849
Norms/influence (parents)	**.095**	**.005**	**.126**	**.004**	.179	.052	.053	.700
Teacher perceptions	-.056	.127	-.017	.728	.188	.137	-.162	.324
Teacher/school careers/events	[AB AC] .006	.837	[AB BC] **.137**	**.001**	[AC BC] **-.268**	**.004**	.003	.980
Explained variance	68.1%		49.8%		48.8%		62.2%	
Unexplained variance (residual)	31.5%		49.7%		36.9%		37.8%	
Unexplained variance (school)	.3%		.5%		14.2%		.0%	

Notes: Missing responses were estimated by expectation-maximisation. Standardised coefficients ('Std. Effect') and significance (p-values; 'Sig. (p)') are reported. Significant coefficients (p < .05) are highlighted in bold. Significant differences in coefficient magnitude (p < .05) across clusters are highlighted in superscript.

Table 35: England 2014/2015 (Years 9, 10, 11): task-level accuracy/bias clusters (four clusters), predicting science intentions (smaller set of predictors)

	Cluster A		Cluster B		Cluster C		Cluster D	
Item/factor	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)	Std. Effect	Sig. (*p*)
Intercept	[AB AC AD] **NA**	**.005**	[AB BC] NA	.448	[AC BC CD] **NA**	**.002**	[AD CD] **NA**	**.016**
Year	[AB AD] 035	.201	[AB] -.046	.185	[CD] .099	.239	[AD CD] -.112	.128
Gender (1=boy)	.006	.817	.003	.912	-.002	.978	-.100	.131
Task-score	[AB] .048	.059	[AB] -.026	.422	[CD] .101	.150	[CD] -.102	.162
Task-confidence	.012	.687	-.041	.213	.027	.691	-.017	.821
Current grade	-.021	.521	-.043	.246	-.064	.426	.074	.428
Self-concept	.063	.069	.071	.091	.056	.564	-.021	.825
Self-efficacy	**.078**	**.022**	**.103**	**.009**	**.190**	**.020**	.123	.130
Interest value	**.123**	**.005**	.069	.156	.024	.832	.219	.062
Utility value	[AB AD] **.565**	**<.001**	[AB BC] **.410**	**<.001**	[BC CD] **.631**	**<.001**	[AD CD] .124	.408
Personal value	**.172**	**<.001**	**.173**	**<.001**	.104	.274	.159	.238
Cost value (absence of)	[AD] .008	.759	[BD] **.091**	**.010**	[CD] .021	.757	[AD BD CD] **-.205**	**.036**
Orientation: mastery	[AD] -.038	.168	[BD] -.004	.923	-.123	.126	[AD BD] **-.263**	**.008**
Orientation: performance	**-.061**	**.020**	-.045	.204	-.101	.151	.091	.301
Perceived control	-.022	.458	[BC] **-.103**	**.007**	[BC] .123	.089	.055	.559
Perceived control (exams)	**-.096**	**.001**	**-.070**	**.041**	-.142	.104	-.092	.239
Study strategy: memorisation	.031	.269	**.102**	**.005**	.139	.054	.169	.095
Anxiety (absence)	.032	.308	.051	.152	.036	.630	.127	.096
Social persuasions (praise)	-.058	.057	**-.101**	**.012**	.048	.573	-.052	.627
Vicarious experiences	-.019	.459	**-.104**	**.003**	**-.157**	**.026**	-.006	.943
Norms/influence (friends)	**-.054**	**.028**	**-.067**	**.047**	.012	.853	-.035	.721
Norms/influence (parents)	**.116**	**<.001**	**.120**	**.004**	.101	.231	.088	.426
Career/events	[AB AC] -.017	.517	[AB BC BD] **.163**	**<.001**	[AC BC] **-.255**	**.001**	[BD] -.119	.214
Explained variance	69.1%		49.4%		55.9%		59.4%	
Unexplained variance (residual)	30.6%		49.9%		39.9%		40.6%	
Unexplained variance (school)	.3%		.7%		4.1%		.0%	

Notes: Missing responses were estimated by expectation-maximisation. Standardised coefficients ('Std. Effect') and significance (p-values; 'Sig. (p)') are reported. Significant coefficients (p < .05) are highlighted in bold. Significant differences in coefficient magnitude (p < .05) across clusters are highlighted in superscript.

Section 9.3: Discussion

In PISA 2006 and in the 2014/2015 survey, the students' science intentions were predicted by different factors in different ways across the various conceptual accuracy/bias groups and across the accuracy/bias clusters from latent-profile analysis, although differences were not apparent for every predictor. Accuracy/bias could then be inferred to be a 'moderator' of some of the associations between the predictors and students' science intentions (Baron & Kenny, 1986).

Some notable differences in coefficient magnitude across the groups and/or clusters appeared to involve the expectancy-value factors. Specifically, considering the conceptual groups for both surveys, those with accurately-high confidence beliefs had the highest predictive association between their perceived utility of science and their science intentions, while students with accurately-low beliefs had the lowest predictive association. Considering the clusters from latent-profile analysis, in PISA 2006, science intentions for the cluster of 'accurately-high' students (cluster D) were more strongly predicted by their perceived utility of science and personal value of science when compared to the cluster of 'accurately-low' students (cluster B), similar to the results from the conceptual groups. A somewhat similar pattern was seen also across the clusters for the 2014/2015 survey when considering the students' utility value. Additionally, utility value had a higher predictive association with intentions for under-confident students than for over-confident students when considering the above-average/below-average conceptual groups in PISA 2006, and when considering the difference-score conceptual groups in the 2014/2015 survey (but only in models with smaller sets of predictors for the 2014/2015 survey, and this pattern was not observed for the regression-residual conceptual groups in PISA 2006).

The students' personal value of science to their identity was predictive for all of the conceptual groups in PISA 2006, but no differences in coefficient magnitude were observed across the above-average/below-average groups. For the regression-residual groups, and also for the clusters, personal value of science had a higher predictive association with intentions for over-confident students compared to under-confident students (cluster C and cluster A). There was little correspondence across the two surveys, however. No statistically-significant differences were apparent for personal value across the clusters in the 2014/2015 survey; for the conceptual groups, however, personal value was not predictive for those with accurately-high confidence, but had a higher predictive magnitude for under-confident students.

The students' interest/enjoyment value of science was predictive for all conceptual groups in PISA 2006, but least strongly predictive for those with accurately-low beliefs (recognising that the coefficients followed this pattern for the regression-residual groups, although were not statistically-significantly different). A similar pattern could be inferred from the clusters, but with fewer statistically-significant differences. Conversely, in the 2014/2015 survey, there were no significant differences of coefficient magnitude for interest across the groups or the clusters. In the 2014/2015 survey, interest value appeared less predictive than other factors even when considering all students (**Section 7.2**), which may explain the absence of differences across the groups or clusters.

The perceived costs associated with studying science were not measured in PISA 2006. The 2014/2015 survey highlighted that the absence of perceived costs associated with higher intentions, but only for the under-confident conceptual group at a small magnitude and with no statistically-significant differences in coefficient magnitude across the groups. The separate factor of science anxiety was generally not predictive for any group or cluster.

Cohering in some respects with earlier research (Dupeyrat, Escribe, Huet, & Régner, 2011; Gonida & Leondari, 2011; Gresham, Lane, MacMillan, Bocian, & Ward, 2000), the over-confident students generally expressed higher attitudes than under-confident students (**Section 8**). Contrary to the relevant hypothesis (**Section 4.2**), however, interest did not appear to be more predictive of intentions for over-confident students compared to under-confident students; for other factors, differences across the methods and the two surveys ensured that the results were less clear.

It remains difficult to contextualise and compare the results against prior studies; relatively little research has explicitly focused on potential differences in decision-making for different students, although some research has plausibly established that different students may find different factors to be more or less relevant to their subject choices (Bøe, 2012). The various results from the two surveys highlight the benefit of considering differences across groups or clusters, rather than considering entire samples, although it perhaps remains unclear what groups or analytical perspectives may be most relevant. Given the focus on students' accuracy/bias, other potential groups were not considered, such as grouping by interest, utility, gender, background, and/or any other factors. For example, students' attainment considered alone has appeared to entail differences in students' decision-making, suggesting that students could simply be analytically grouped by attainment (Brown, Brown, & Bibby, 2008). Similarly, further research has highlighted that girls and boys may consider different factors and/or the same factors in different ways when making their studying choices, although potentially contrasting results can be seen across different studies (Bøe, 2012; Bosker & Dekkers, 1994; Crombie, et al., 2005; Department for Children, Schools and Families, 2009).

The varying predictive magnitudes of the expectancy-value factors across the groups and clusters may potentially provide insight to further link motivational theories and analytical perspectives with students' choices. In contemporary research, the expectancy-value model of students' choices proposes a general structure and key factors, including specific aspects of 'value' that are theorised to reciprocally associate with each other and with students' confidence considered as expectations of success (akin to self-efficacy beliefs) (Eccles, 2009; Eccles, et al., 1983; Wigfield & Eccles, 2000). Sometimes the measures of value have been aggregated together rather than being considered separately (Chow, Eccles, & Salmela-Aro, 2012; Wang, 2012). Essentially, the contemporary expectancy-value model assumes that high interest, and/or utility, and/or other factors, may entail higher intentions. High beliefs in any one factor may be sufficient for high intentions, for example where higher perceived utility may cover for lower interest. Most predictive modelling implicitly follows a similar assumption of independence; any one factor considered alone is sufficient to predict high intentions. Alternately, historical models of students' choices often assumed that motivations towards choices or actions followed from some form of complex aggregation or interactions between the various modelled factors, such as choices following from motives multiplied by expectancies multiplied by incentives (Atkinson, 1957, 1964; Wigfield & Eccles, 1992). Adapting the idea of 'interactions' from historical models of motivation or choices may suggest that high beliefs across all (or some) attitudes are necessary for high intentions, and/or that the interaction between someone holding both high interest and high utility (for example) might entail that they hold greatly higher intentions.

Some prior research has explored interactions between various factors. For example, across all countries in PISA 2006, science intentions were predicted by the students' self-concept beliefs, their interest, and (at a small magnitude) the interaction between self-concept and interest (Nagengast, et al., 2011). Similarly, for secondary school students in Germany, self-concept beliefs, aspects of value, and their interaction, all predicted higher attainment (Guo, et al., 2016; Trautwein, et al., 2012). However, it remains unclear which particular attitudes would interact with which particular factors when predicting different outcomes. For example, prior research either omitted students' perceived utility value (Nagengast, et al., 2011) or found that it had little predictive association with attainment and hence was less relevant within interactions with other factors (Guo, et al., 2016). Such studies perhaps establish the plausibility of a general idea, rather than entail definitive models to be applied in other contexts.

The results from PISA 2006 and the 2014/2015 survey suggest that students' perceived utility of science may have an increased influence on their intentions for those with accurately-high beliefs (high confidence and high attainment), which perhaps suggests that interactions between factors are plausible. Future research may need to explore whether and how utility value and perhaps personal value (and/or other factors) interact with students' other attitudes, attainment, and/or confidence when predicting their intentions. However, researchers may need to remain mindful of what factors intend to measure and represent (such as whether utility value, representing 'indirect value', may also have elements of 'direct value' for some students), and how interactions between factors would be interpreted to help understand students' intentions and choices. As predictive models become increasingly abstracted, elements such as interactions may become harder to link to students' potential processes of decision-making, but the area could perhaps be alternately explained or explored via considering clusters of students. Cluster membership can be considered as another perspective that helps explain the overall patterns of association seen across an entire sample (Collins & Lanza, 2010). In PISA 2006 for example (**Section 8.1.3**), the intention/attitude cluster with the highest, above-average, attitudes exhibited even higher perceived utility of science (and to a lesser degree, science intentions) than their other attitudes; modelling interactions between utility and other attitudes when predicting intentions across an entire sample might essentially help account for such (smaller numbers of) students. The area could alternately/additionally then be explored by attempting to consider what predicts cluster membership, and/or students' attitudes themselves.

The results highlighted various other points of note, when considering further factors. Prior studies have highlighted that students' intentions may be influenced by parental or other encouragement in various ways (Buschor, Berweger, Frei, & Kappler, 2014; Mujtaba & Reiss, 2014; Sjaastad, 2012). The 2014/2015 survey results extended on such earlier studies and highlighted that for the conceptual group of over-confident students, in comparison to other groups, science intentions were predicted relatively more by perceived norms/influences from the students' parents. Further research may be necessary to explore whether any influences from other people are perceived by students as support or as expectations to be met. If over-confident students are especially encouraged to study science, then they may need further academic support in order to ensure that they can gain any pre-requisite attainment for studying science further. It also remains unclear whether encouragement might, in some cases, lead to over-confidence (and/or whether over-confidence is partially rationalised via perceived expectations).

In the 2014/2015 survey, across all students, the students' age (academic year) had no predictive association with their intentions, once the students' attitudes were modelled. However, when considered for the conceptual accuracy/bias groups, older students were predicted to have higher science intentions, accounting for their attitudes and background, but only for those with accurately-high beliefs. It is perhaps plausible to infer that some students may refine their intentions over time and become increasingly sure of particular aspirations (Cleaves, 2005).

In England, declining attitudes to science as students grow older has often been considered as a cause for concern (Osborne, Simon, & Collins, 2003). In other countries, changes in students' attitudes over time have often involved largely-positive views declining to slightly-positive or neutral views when considered across various academic subjects (Archambault, Eccles, & Vida, 2010; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Nagy, et al., 2010). Similarly, different patterns of change have been observed for different groups of students, for example with some attitudes remaining relatively stable over time for some students or some attitudes declining but remaining high or moderate for other students (Archambault, Eccles, & Vida, 2010; Köller, Baumert, & Schnabel, 2001; Musu-Gillette, Wigfield, Harring, & Eccles, 2015). Further research may then need to identify groups of students in England and explore in more detail how attitudes decrease or increase over time.

Both surveys highlighted that the reported provision of careers information or guidance was generally not predictive or was less predictive for the conceptual groups with accurately-high beliefs, although the magnitudes varied across the two surveys for the other groups. Considering the clusters, careers information was more predictive of science intentions for the 'accurately-low' cluster than for the 'under-confident' and 'accurately-high' clusters in the 2014/2015 survey. A similar pattern of coefficient magnitudes was seen for the groups and clusters in the PISA 2006 survey, although differences were only statistically-significant across the regression-residual groups.

Promoting science through careers information and guidance has often been emphasised within science education, especially as the National Curriculum has not explicitly conveyed a requirement for teachers to explain the careers available within science and the careers that may be facilitated by science skills or qualifications (Archer & DeWitt, 2015; Department for Education, 2013, 2014; Osborne & Dillon, 2008; Royal Society, 2014). The National Curriculum does not appear to cover a need to explain careers information for any subject, however, perhaps due to the need for careers information being addressed through other statutory guidance, which does highlight a need to convey that many careers require high mathematics and science knowledge (Department for Education, 2015). It is possible that teachers may discuss careers within science or other lessons regardless, in addition to any provision from specialised careers staff and external sources. While the nature and provision of careers advice at school has indeed been associated with students studying physical science subjects, the extent or mechanism perhaps remains unclear (Bennett, Lubben, & Hampden-Thompson, 2013; Blenkinsop, McCrone, Wade, & Morris, 2006). For example, interviews with university students in Denmark highlighted that they had considered their choices as a personal responsibility and that advice from school careers staff had not been frequently sought (Holmegaard, Ulriksen, & Madsen, 2014). In many research studies that highlight the influence of teachers, it also remains unclear whether this involved formal or informal careers guidance and/or other forms of support (Aschbacher, Li, & Roth, 2010; Maltese, Melki, & Wiebke, 2014; Sjaastad, 2012). Given the results from PISA 2006 and the 2014/2015 survey, it perhaps remains important to consider and explore whether different students find careers advice more or less helpful, and how careers advice might influence students' intentions (potentially directly) and/or influence students' attitudes (such as someone's perceived utility of science, and hence indirectly influence their intentions).

Section 9.3.1: Conclusions

Research question three: did students with different degrees of confidence accuracy/bias consider their science intentions in different ways?

Fundamentally, the results from the PISA 2006 survey and the 2014/2015 survey highlighted that science intentions were indeed predicted by different factors and/or in different ways, especially for factors such as the students' utility value and personal value of science, although with some variation across the methods and surveys. These highlighted that attitudes such as utility value could be more predictive of intentions for students with accurately-high confidence beliefs; depending on the methods, utility value could be more predictive of intentions for under-confident students than over-confident students, and personal value could be more predictive for over-confident students than under-confident students, or that no differences could be observed for these factors. These inconsistencies highlighted that further research would be needed to clarify the area. Nevertheless, the results broadly highlighted that the impact of under-confidence and over-confidence may be more complex than previously assumed within prior studies (Bouffard & Narciss, 2011): confidence biases may associate with lower or higher attitudes, and also with students considering their choices in different ways.

Section 10: General discussion

The preceding results are now summarised, broadly focusing on highlighting the new knowledge and insights gained (**Section 10.1**). As in any research, various limitations may determine how meaningful the results are, and may raise implications for future research (**Section 10.2**). Finally, implications for practice and policy are suggested (**Section 10.3**) and conclusions are drawn (**Section 10.4**).

Section 10.1: New knowledge and insights gained

An increased understanding of students' intentions to study science at upper-secondary school may help wider consideration of varying or imbalanced progression (**Section 1**). Prior research into science education in England has not extensively considered the potential impact of students' confidence accuracy/bias (**Section 2**), although such bias may associate with students' attitudes and motivational beliefs (**Section 3**). The research presented in this thesis broadly aimed to enhance understanding of the area by applying confidence accuracy/bias as an analytical perspective (**Section 4**). The research considered two surveys (**Section 5**), PISA 2006 and a new survey undertaken in 2014/2015, to address methodological limitations in one set of data with strengths in the other, and so that any consistent results across varying approaches would enhance their plausibility. Emerging methods such as identifying clusters of students with different degrees of confidence accuracy/bias via latent-profile analysis were also applied (**Section 6**), which appeared not to have been previously used to consider confidence accuracy/bias in England.

The first research question examined which attitudes and motivational beliefs (including expressions of confidence) were the most relevant influences on students' science intentions (**Section 7**). Across both surveys, students' perceived utility of science and personal value of science were most strongly predictive of science intentions, together with other factors. This provided a new insight in itself, since the personal value of science to someone's identity has generally been explored through qualitative perspectives where the relative impact compared to other factors could not easily be determined (Archer, et al., 2010; Aschbacher, Li, & Roth, 2010; Carlone & Johnson, 2007). On a wider level, the subsequent formation of clusters of students, based on their intentions and attitudes (specifically, including the key predictors of intentions such as personal value), highlighted some coherence across these factors for students within each cluster. Accordingly, science education may benefit from exploring how ideas of science identity can be considered and developed further within quantitative approaches, such as extending ‘personal value’ factors to consider more dimensions of identity and/or considering the structure, coherence, or associations between multiple factors.

Across both surveys, the indicators of attainment appeared to mediate the predictive association between the number of books at home and students’ intentions, while the indicators of confidence appeared to mediate the association between gender and intentions. The predictive association between intentions and parents working within science appeared to be mediated by attainment in PISA 2006 and by utility value in the 2014/2015 survey. Additionally, in the 2014/2015 survey, utility value and norms/influences from parents formed separate factors and were both independent predictors of students’ science intentions. These results provide another wider insight, suggesting that aggregating indicators of parents working within science, utility value, norms/influences from parents, and other aspects as ‘science capital’ should perhaps be considered cautiously (Archer, Dawson, DeWitt, Seakins, & Wong, 2015). Ideas such as science identity and science capital could perhaps be developed further within social-cognitive models (Eccles, 2009; Wigfield & Eccles, 2000), separating potential antecedents, such as parental attitudes or parents working within science, from students’ attitudes and beliefs, while further considering and clarifying what ‘identity’ and ‘capital’ may entail.

Across both surveys, the students’ intentions were predicted more by utility value than interest value; prior research in England has similarly highlighted the importance of utility value (Mujtaba & Reiss, 2014; Vidal Rodeiro, 2007) or highlighted the importance of interest (Bates, Pollard, Usher, & Oakley, 2009; Mellors-Bourne, Connor, & Jackson, 2011). Utility value aims to measure students’ perceptions of the indirect value of studying science (Eccles, 2009), such as to gain skills that could be employed in many areas, but this may be inseparable from some aspects of direct value: for those who intend a career in science, utility value may potentially reflect the recognition that studying science at various stages of education is unavoidably necessary for a science career. On a wider level, it may be useful to consider how studying science may be directly and/or indirectly valued, and how factors are conceptualised and interpreted (Archer, Dawson, DeWitt, Seakins, & Wong, 2015; Claussen & Osborne, 2013; Williams & Choudry, 2016). Concurrently, it may be beneficial to identify antecedents of attitudes such as utility value.

Students' confidence expressed as self-concept beliefs was a strong predictor of their science intentions in the PISA 2006 survey, while confidence expressed as self-efficacy beliefs (contextualised as confidence in their expected grades at GCSE/A-Level) was a strong predictor in the 2014/2015 survey. Indicators of accuracy/bias appeared to be directly predictive of science intentions if the underlying expression of confidence was also predictive, as in PISA 2006. Measures of task-level accuracy/bias were ultimately not predictive of science intentions in the 2014/2015 survey, however, when also accounting for the students' various attitudes and beliefs.

Indicators of confidence, attainment, and accuracy/bias could not be included together within predictive modelling, since any one of these could be perfectly predicted by any two of the others, given the calculation approaches; the impact of accuracy/bias was then unavoidably less clear when considering all students together (**Section 7**). The impact of accuracy/bias was then considered via grouping or clustering students, to determine whether students with different degrees of accuracy/bias nevertheless expressed different science intentions and attitudes (the second research question, **Section 8**), and whether such students might consider their choices in different ways (the third research question, **Section 9**). Both of these research questions were positively answered by the results from the two surveys: students with different degrees of accuracy/bias did express different science intentions and attitudes, and could be inferred to consider their choices in different ways, which was another new finding.

The application of multiple approaches to classify students also highlighted another insight: emergent clusters of students could be broadly described as showing tendencies towards under-confidence, accurately-low confidence, accurately-high confidence, or over-confidence. This validated the use of conceptual accuracy/bias groups (formed from particular conceptual definitions of ‘under-confidence’ and ‘over-confidence’, partially depending on a methodological approach or technique), especially since similar results from emergent clusters have been observed outside of England for mathematics (Sáinz & Upadyaya, 2016). While the conceptual groups and the emergent clusters embodied different proportions of students with different profiles (so that descriptive labels such as ‘under-confident’ would entail different magnitudes of confidence and attainment in a group or in a cluster), some similarities in results were still observed across the approaches.

Students with accurately-high confidence generally reported the highest intentions and attitudes, across both surveys and for the conceptual accuracy/bias groups and the accuracy/bias clusters (**Section 8**). The results from PISA 2006 showed that under-confident students exhibited lower intentions and attitudes than over-confident students across the conceptual groups and emergent clusters. The results from the 2014/2015 survey provided partial support: the under-confident and over-confident conceptual accuracy/bias groups reported similar science intentions and self-efficacy, but under-confident students reported lower attitudes including interest, utility, and personal value, which were important predictors of science intentions (**Section 7**). Conversely, however, the accuracy/bias clusters in the 2014/2015 survey showed that those exhibiting over-confidence reported lower intentions and attitudes than those exhibiting under-confidence. Groups/clusters with ‘accurately-high’ confidence could also or conversely appear over-confident (to varying extents), depending on the considered indicators, perhaps suggesting the contextual or conditional nature of any benefits or detriments of confidence biases, such as requiring some level of underlying attainment and/or confidence.

On a wider level, the results then highlighted that results could partially depend on particular analytical approaches, which may be problematic if a study only considers one approach; while conceptual groups reflected tendencies towards accuracy/bias that were indeed observed in emergent clusters (to varying extents), the specific profiles of conceptual groups and clusters unavoidably differed. Across the two surveys, the results broadly cohered with earlier research that associated higher interest with over-confidence rather than under-confidence (Gonida & Leondari, 2011), but contrasted when considering some other motivational beliefs. Earlier studies have highlighted that over-confident students have reported higher out-performing orientations than under-confident students (Dupeyrat, Escribe, Huet, & Régner, 2011; Gonida & Leondari, 2011), for example. However, no difference was observed across the under-confident and over-confident accuracy/bias groups in the 2014/2015 survey, while the under-confident accuracy/bias cluster exhibited higher beliefs than the over-confident cluster regarding their orientations to master learning and to out-perform others in science. Given the differences in profiles (magnitudes of attainment and confidence) across the groups and clusters, further research may need to consider whether such motivational beliefs follow more from confidence, attainment, and/or accuracy/bias.

Concurrently (**Section 9**), across both surveys, despite some variability across the groups and clusters, it can be plausibly inferred that the students' perceived utility value was more predictive of intentions for those with accurately-high confidence compared to other groups and clusters. Careers information appeared to predictively associate with intentions to a relatively greater extent for those with accurately-low confidence. Depending on the method/approach, utility value could be more predictive of intentions for under-confident students than over-confident students, and personal value could be more predictive for over-confident students than under-confident students, or that no differences could be observed (although the magnitudes of coefficients still broadly followed these patterns when statistically-significant differences were not confirmed). Now that differences across methods and surveys have been revealed, further research would be beneficial to clarify the area.

On a theoretical level, higher confidence, contextualised as self-efficacy beliefs or someone's expected capability in future tasks and activities, has been theorised to be motivational and facilitate people to surpass their normal performance, while low confidence beliefs may be limiting and ensure that some actions are not even attempted (Bandura, 1997). This implies that under-confidence may be detrimental while over-confidence may be beneficial. Alternately, the idea of self-regulation assumes that accurate beliefs are integral to personal well-being and functioning (Butler & Winne, 1995). This implies that under-confidence and over-confidence may be equally detrimental.

Much research has focused on this potential duality between the underlying social-cognitive theory (Bandura, 1997) and ideas of self-regulation (Butler & Winne, 1995), for example considering whether over-confidence may be beneficial or detrimental (Bouffard & Narciss, 2011). These areas may not necessarily be mutually exclusive, as social-cognitive theory has been contextualised within cyclical model of self-regulation (Bandura, 1989, 2001), and self-regulation is often conceptualised within a social-cognitive framework (Zimmerman, 2000). It remains possible that self-regulation may ensure general tendencies towards accuracy, while higher confidence may be motivational. The various results broadly support this conclusion: essentially, magnitudes of confidence bias were less extreme in the emergent clusters of students than assumed by the various conceptual groups; the students' attitudes were broadly proportional to their confidence and those with the highest confidence and attainment also exhibited the highest accuracy.

Section 10.2: Limitations and implications to further research

Through considering the PISA 2006 survey (considering implicit subject-level self-concept accuracy/bias) and through applying a new survey in 2014/2015 (considering explicit task-level confidence accuracy/bias), the overall approach aimed to address methodological limitations in one survey with strengths in the other. Some similar results were observed across both surveys, regardless of the inherent differences in students, measurement, and approaches to consider accuracy/bias, which enhanced the overall plausibility of these findings. Different results were also observed, however. As in any research, various issues may determine how meaningful the results are and may suggest areas for further exploration.

The two surveys applied some comparability in measurement but also some differences (**Section 5** and **Appendix 1**). The 2014/2015 survey was intended to be comparable across both PISA and TIMSS in order to facilitate emergent analysis or wider explorations; the questionnaire development also considered subsequent surveys such as PISA 2012 in addition to PISA 2006 in case of potential refinements or developments in measurement having occurred over time (OECD, 2009a, 2013). The development process theoretically increased construct/factor validity, where the measured items then reflected established operationalisations; similarly, content validity, considered as the scope of the measurement items, was also theoretically increased through including items covering multiple themes or dimensions from across different prior surveys. The similar results, for example where utility value and personal value were the strongest predictors of intentions in both surveys (**Section 7**), highlighted the predictive/criterion validity of these factors (Cohen, Manion, & Morrison, 2007; Messick, 1995). Essentially, the 2014/2015 survey highlighted that different operationalisations of factors (e.g. personal value) could be strongly grounded in particular theoretical perspectives (e.g. Carlone & Johnson, 2007), relatively briefly measured, show acceptable internal consistency and reliability, and broadly reveal similar findings to those observed in PISA.

The research design was not necessarily ideal, however. Construct/factor validity could be maximised by using items from only one source, although this would assume that the selected source covers all necessary dimensions of the underlying idea. Given that attitudes and beliefs can be potentially defined in different ways, it remains difficult to determine what items/dimensions are necessary and/or sufficient in measurement, and hence it remains difficult to definitively ensure content validity (Murphy & Alexander, 2000; Wigfield & Cambria, 2010). Construct/factor validity also depends on historical and/or contemporary interpretations of any underlying ideas, yet personal value or ‘science identity’ can still be considered in various ways (e.g. Archer, et al., 2010; Aschbacher, Li, & Roth, 2010; Carlone & Johnson, 2007). Further research may need to apply qualitative approaches or facilitate free-text responses from students in order to explore how they consider or define discrete dimensions within their interest in science, perceived utility of science, personal value of science, and other areas. This would help ensure ecological validity, given that students may find various experiences or aspects of life to be relevant to their interest or identity (and their other attitudes and beliefs) regardless of what contemporary theories may propose, and also help ensure face validity, given that students may have different ‘intuitive’ interpretations of items (and/or wider ideas) compared to any intended meanings.

Differences across the surveys were observed for factors such as interest value, highlighting that predictive/criterion validity could vary. It remains unclear whether item-level measurement similarity/difference influenced any differences in results (e.g. personal value was measured with greater difference across the surveys yet showed similar results; **Section 7**). Aggregating across Years 9, 10, and 11 in the 2014/2015 survey may have ensured that some effects were potentially harder to observe if they varied by age. For example, the similar predictive associations between personal value and intentions across both surveys may suggest that the underlying ‘effect’ was independent of age; the different predictive associations between interest value and intentions across the surveys may suggest that the effect was potentially obscured in the 2014/2015 survey by differences across ages (where different predictive associations depending on students’ ages may have potentially reduced the overall observed association). With relatively few students and different numbers in each year group, the 2014/2015 survey could not feasibly explore differences by age, and further research would be necessary to explore this area. Alternately, differences across surveys may unavoidably occur, given different cohorts of students surveyed at different times, and/or some results may have been specific to the particular sample of students surveyed in 2014/2015 who were not necessarily generalizable to the wider population of students across England.

Aspects of measurement were also relevant to considering confidence accuracy/bias. Fundamentally, indicators of accuracy/bias formed through comparisons of confidence and attainment may inherently involve uncertainty (**Section 3.4**), which was especially relevant for the PISA 2006 data. Comparing students’ self-concept beliefs against different indicators of attainment may produce different results; any comparisons may be artificial, and cannot confirm or entail that students self-evaluate in the same way. Considering paired tasks and confidence-ratings in the 2014/2015 survey theoretically measured self-reflective accuracy/bias with stronger validity, but task-level indicators may have lower generalisation to students' subject-level attitudes and beliefs (Bandura, 1997; Bong, 1997; Pajares & Miller, 1995), which appeared to be highlighted in the analysis (**Section 7**). Nevertheless, the similarities across both surveys (broadly seen in **Section 8** and **Section 9**) supported the plausibility of exploring accuracy/bias even when approximated via PISA data.

When grouping or clustering the students in the 2014/2015 survey on their task-confidence and task-scores, their self-concept beliefs and current grades followed broadly proportional patterns, suggesting that, in general terms, the same tendency towards accuracy/bias might be observed on the task and subject levels. The calculated indicators of accuracy/bias showed low to moderate correlations across the task and subject levels, however, although such results may be less clear due to accuracy/bias being considered explicitly on the task level (via paired tasks and confidence-ratings) and implicitly on the subject level (via self-concept beliefs and subject grades, where various other measures of attainment may be relevant). Even considering prior research, it still remains relatively unclear whether students' confidence accuracy/bias generalises from the task-level to the subject-level, and/or whether either or both can reveal tendencies that are specific to subjects or tendencies that are more generalised. For example, for secondary school students in Greece, there appeared to be no clear pattern of tendencies towards similar biases across mathematics and languages (Gonida & Leondari, 2011). Studies with undergraduate students in the United States have highlighted some similarities of accuracy/bias across domains or areas, although these have not reflected academic subjects and may have less contextual relevance to secondary school students (Gutierrez, Schraw, Kuch, & Richmond, 2016; Schraw, Dunkle, Bendixen, & Roedel, 1995; West & Stanovich, 1997). However, studies with primary and secondary school students, and university students, have suggested some stability of accuracy/biases over time (Bouffard, Vezeau, Roy, & Lengelé, 2011; Kelemen, Frost, & Weaver, 2000; Rytkönen, Aunola, & Nurmi, 2007). Further research may be useful to explore such areas.

Exploring confidence accuracy/bias in general may be informed by further research into how students form their confidence beliefs, which may benefit from more exploratory or qualitative approaches (Butz & Usher, 2015; Zeldin & Pajares, 2000; Zeldin, Britner, & Pajares, 2008). The 2014/2015 survey started to explore and extend such areas through asking students, for example, what grade entailed ‘being good’ at science, which showed less variation across the accuracy/bias groups and clusters. Accordingly, it may be plausible that students have relatively similar ideas about what entails good attainment, which can be explicitly measured so that self-concept beliefs (measured through agreement or disagreement with, for example, ‘I usually do well in science’, ‘I have always been good at science’, and ‘I get good grades in science’) and other indicators of confidence can then be compared against these ideas (with increased validity). However, given the focus on various other groups and clusters, and for brevity, this area was left for future research. Further longitudinal research would also be necessary to explore the accuracy/bias of students’ self-efficacy expressed through expected grades.

Similarly, and on a wider level, further research may need to continue to apply more qualitative studies where students can express their own reasons for their subject choices, and also define areas such as utility value and personal value in their own way. Ideally, such findings could then also aid the refinement of measurement items (Clark & Watson, 1995; Messick, 1995; Simms, 2008); for example, the conception and measurement of areas such as utility value may need to be calibrated against students’ conceptions of the area. Qualitative studies have often highlighted that students considered interest and enjoyment more relevant to their choices than utility, which may highlight potential differences between qualitative and quantitative results, highlighting the benefit to consider and integrate findings from across both approaches (Holmegaard, Ulriksen, & Madsen, 2014; Jensen & Henriksen, 2015).

Quantitative approaches allow independent associations between factors and expressed intentions to be revealed, but results from predictive models do not necessarily mean that students form their intentions or make their choices in the same way as suggested by the various significant predictors. Students may consider their choices and intentions in different ways, perhaps depending on their gender (Bosker & Dekkers, 1994; Crombie, et al., 2005), and/or their background (Archer, et al., 2012, 2013; Archer, DeWitt, & Wong, 2014), and/or their focus on particular subjects (Bøe, 2012). A potential difference in decision-making due to confidence accuracy/bias was plausible, but further research may need to determine which analytical perspectives are most relevant or explanatory.

Analytical models are simplified representations of reality, designed to accomplish particular purposes. The particular purpose may explicitly influence results by determining what research questions are asked and hence what results are considered to be meaningful or not. The various simplifications within analysis may implicitly influence results, such as through factors being included or omitted, or through particular methodological approaches being applied. Specifically, the research covered in this thesis applied confidence accuracy/bias as a broad analytical perspective, in order to increase understanding of students' intentions to study science. Various other perspectives are possible, including those based on more observable characteristics such as gender and background/ethnicity (Carlone & Johnson, 2007; Murphy & Whitelegg, 2006), or less observable areas including social and cultural aspects such as expectations and stereotypes (Cundiff, Vescio, Loken, & Lo, 2013; Forgasz, Leder, & Kloosterman, 2004), and ideas of general advantage and disadvantage transferred in part through resources or 'capital' (Archer, Dawson, DeWitt, Seakins, & Wong, 2015; Claussen & Osborne, 2013). Ultimately, the impact of students' confidence accuracy/bias may be relatively small, for example when considering the magnitude of differences in attitudes across the various groups and clusters from PISA 2006 and the 2014/2015 survey; it is possible that other perspectives might reveal larger differences or effects.

Predictive modelling, regardless of the level of statistical sophistication, relies on underlying assumptions such as linear associations. Practical contexts may involve complex combinations of linearity but also thresholds: at some particular magnitude of interest, perceived utility, and/or any other factors, someone might chose to study science rather than not study science. Any thresholds may also potentially vary for different students. Prior research broadly suggests that various thresholds may be relevant, most obviously the expected or actual grades that may be required by schools as pre-requisites for A-Level study (Brown, Brown, & Bibby, 2008; Department for Children, Schools and Families, 2009). The analytical approach indirectly addressed this area through applying latent-profile analysis to identify clusters of students based on their science intentions and key predicting factors; essentially, various magnitudes of these factors (perhaps in combination) might entail thresholds for membership of one cluster compared to another. Across both surveys, only the smallest clusters with consistently highly above-average attitudes exhibited agreement to study science further, while the other clusters generally exhibited some degree of disagreement or ambivalence. Further research could beneficially explore this area through predictive modelling, perhaps through considering intentions as ordered categories, and/or through more complex approaches. It may also be beneficial to consider what predicts cluster membership, and/or what might entail movement from one cluster to another.

The two surveys both considered science as a holistic area. While this reflected the National Curriculum and simplified the measurement of students' attitudes and beliefs, students may have varying views across biology, chemistry, physics, and other relevant areas (Hardy, 2014; Jansen, Schroeders, & Lüdtke, 2014). Ultimately, students would also need to select biology, chemistry, physics, and/or other specific subjects for A-Level and/or university study. The results may then provide a plausible overview, but may be less meaningful to help understand students' actual decisions.

Studying one subject at upper-secondary school or university in England may entail not studying another subject or subjects. Considered in general terms, it may be difficult to gain a comprehensive understanding of students' intentions and choices to study science without also considering what other subjects might be favoured instead of science and why. For example, it may be difficult to understand why someone with high interest and perceived utility for science might not study science, without considering whether they had even higher interest in another subject; relative differences in attitudes across subjects may then be equally or more relevant than considering attitudes in one subject alone. Research has indeed suggested that some students may study other subjects (and not science) due to favouring other subjects rather than necessarily disfavouring science (Cleaves, 2005). Prior research has also suggested that clusters of students may prefer sciences, humanities, or other areas, although the particular patterns and profiles of students remains somewhat unclear (Institution of Mechanical Engineers, 2014). This area could be addressed through future research considering limited arrays of items/factors but covering students' attitudes and beliefs across multiple academic areas.

Various other analytical aspects may be relevant limitations to the thesis. For example, the analysis only considered cluster solutions with relatively small numbers of clusters. Higher number of clusters essentially entailed fewer numbers of students within each cluster, ensuring that predictive modelling per cluster and multiple comparisons were less feasible. An alternate approach might involve considering higher numbers of clusters, perhaps ignoring those with very few numbers of students (i.e. potentially 'outliers' due to uncertainty or unreliability); such decisions might become arbitrary, however, and there may be no easy resolution.

Additionally, latent-profile analysis identifies discrete clusters of students, given the various indicators (Collins & Lanza, 2010; Everitt, Landau, Leese, & Stahl, 2011). The meaning of the cluster profiles then needs to be interpreted, and the clusters may not necessarily reflect what they were intended to reflect. For example, the analysis identified clusters of students in England in PISA 2006 given their self-concept beliefs and task-scores, to implicitly form clusters with different degrees of self-concept accuracy/bias; one cluster of students was identified with an essentially unvarying magnitude of self-concept and with lower but more variable task-scores (while other clusters showed within-cluster variation for both indicators). The analysis simply revealed emergent clusters within the data, and the students may have expressed those particular self-concept beliefs for many reasons. For example, the students may have essentially agreed to every self-concept item with less reflection on the content (perhaps given that the items were on the last page of the questionnaire); the result may have then reflected some form of response style or tendency rather than reflecting self-concept accuracy/bias. Conversely, it cannot necessarily be determined that particular response styles were present since the students expressed variability in their other attitudes. The results nevertheless highlight that the appearances of accuracy/bias may perhaps follow from low self-reflection or from other aspects that may not necessarily be easily determined or isolated.

It remains uncertain whether students' reported attitudes and beliefs in surveys reflect their actual attitudes and beliefs, although this is an issue for any research study. While anything that reduces the correspondence between indicators of confidence and attainment may give the appearance of under-confidence or over-confidence, it remains unclear whether and to what extent random variation or other measurement aspects may influence appearances of accuracy/bias. Studies have variously proposed that confidence biases are explained or are not explained by random variation (Brenner, 2000; Budescu, Wallsten, & Au, 1997; Erev, Wallsten, & Budescu, 1994; Juslin, Winman, & Olsson, 2000; Merkle, Sieck, & van Zandt, 2008). It perhaps remains pragmatic to assume that indicators of accuracy/bias will unavoidably involve some degree of imprecision and uncertainty, but still measure a meaningful idea.

Assuming that some students may potentially agree to items regardless of their content and/or may respond differently to positively-phrased or negatively-phrased items (Cronbach, 1950; Paulhus, 1991; Weems, Onwuegbuzie, & Collins, 2006; Weems, Onwuegbuzie, Schreiber, & Eggers, 2003), it may be beneficial for further research to explore any associations between response styles and confidence accuracy/bias. For example, studies have begun to explore over-confidence defined as students expressing familiarity with or knowledge of non-existent concepts, which combines ideas of an acquiescence response style with a confidence bias, and which has associated with other indicators of over-confidence biases (Paulhus & Dubois, 2014; Paulhus, Harms, Bruce, & Lysy, 2003). Nevertheless, it remains unclear whether response styles such as acquiescence are prevalent or uncommon. Prior research on PISA 2006 has suggested, for example, that the majority of students in Germany did not appear to respond with a particular style such as only selecting the extreme response categories (Wetzel, Carstensen, & Böhnke, 2013). It remains difficult to conclusively consider the area without carefully designed questionnaires, for example containing similar numbers of positively-phrased and negatively-phrased items for all factors, which may be difficult to practically implement.

On a wider level, confidence biases may potentially follow from particular motivations, such as to self-enhance through maximising positive beliefs or to self-protect through minimising negative beliefs (Alicke & Sedikides, 2009; Sedikides & Alicke, 2012). Alternately, confidence biases may potentially follow from low self-reflection, or from students forming their beliefs against subjective criteria or being influenced by other factors, without any particular underlying enhancement or protection motivation. Confidence biases may also associate with traditional characterisations or measures of personality, although results have varied within psychological studies of university students (Anderson, Brion, Moore, & Kennedy, 2012; Schaefer, Williams, Goodie, & Campbell, 2004). These areas may nevertheless suggest further factors to be considered in future research, especially to help understand why particular confidence accuracy/bias tendencies might form, and hence how they might plausibly be amended.

Section 10.3: Practical and policy implications

Wider policy recommendations within science education have recently included focusing on a need for inspirational curricula, which emphasise practical work and problem-solving, and have highlighted that increased science careers guidance should be given, that students should study science throughout upper-secondary education to help broadly increase students' skills, and that further specialised teachers are needed (Osborne & Dillon, 2008; Royal Society, 2014). The various insights from this thesis do not easily lead to policy implications on such a wide scale, and perhaps fundamentally suggest a more cautious approach: without understanding how different students make their choices in different ways, broad changes to practices may not necessarily achieve any intended increases in the numbers of students studying science. Differences in students' decision making may include varying influences (e.g. while the perceived utility of science was predictive of intentions for all students, it was most strongly predictive for those with accurately-high confidence), and varying scope (e.g. the reported provision of careers information was predictive of intentions for some groups and clusters but not for others). In order to increase numbers of students studying science, researchers and teachers may need to consider target groups (e.g. under-confident students who might already have sufficient attainment) and consider how they make their choices in more detail.

With reference to existing policy recommendations such as forming and applying inspirational curricula (Royal Society, 2014), the various results and insights from the thesis confirm that attitudes such as students' interest associate with students' intentions, but that students' perceptions of their teachers and reported frequency of experiencing different teaching approaches had limited associations with intentions. Teaching approaches and classroom experiences may influence students' interest and other attitudes, but research may need to determine to what extent and whether other influences might be more relevant (Abrahams, 2009; Hampden-Thompson & Bennett, 2013; Wang, 2012). Similarly, the various results suggest that careers guidance may not necessarily have large or direct impacts on students' intentions to study science, on average, but may be more relevant to students' with accurately-low confidence (who generally expressed below-average levels of background resources and parental education) and so may potentially help equality. While international comparisons were not a feature of the thesis, without extensive changes to upper-secondary education in England, compulsory study of science would simply entail that students cannot study one other subject; alternately, changing the educational structure to the one used in Sweden (with broad but relatively few programmes of specialism that still include compulsory areas such as general science across all programmes) or Finland (with multiple compulsory basic areas including sciences and multiple optional areas of further specialisation) may not be easily or swiftly undertaken (EACEA, 2011; Eurydice, 2016). Essentially, some existing policy recommendations can be broadly supported through the results, while others may need to be clarified and considered further.

Intuitively, science skills may be less relevant without the confidence to apply them; similarly, over-confidence may be problematic if someone ultimately lacks the attainment necessary to progress within science. The results from the PISA 2006 survey and the 2014/2015 survey broadly supported such assumptions through highlighting the relevance of students' confidence to their science intentions; the final impact may also be complex due to science intentions being predicted differently for students with different degrees of accuracy/bias. Practically, it may be beneficial to further consider how different students make different choices, which may also entail considering multiple perspectives such as students' background, resources, and attitudes (Archer, Dawson, DeWitt, Seakins, & Wong, 2015; Bøe, 2012; Carlone & Johnson, 2007; Murphy & Whitelegg, 2006). Additionally, students may differ even within particular classifications (e.g. girls may have various orientations towards science and distinct clusters may be apparent), which may entail some unavoidable degree of complexity when considering which perspective might be most informative.

The association between students' science intentions and their self-efficacy expressed as expected grades in the 2014/2015 survey may suggest that teachers could help provide feedback to ensure that students have realistic expectations. Similarly, the association between students' intentions and their self-concept beliefs in the PISA 2006 survey may suggest that teachers need to be aware of potential under-confidence or over-confidence. However, various studies have highlighted that teachers' beliefs of their students' attainment may not necessarily correspond to the students' attainment (Harlen, 2005; Südkamp, Kaiser, & Möller, 2012), and that teachers may not easily or accurately estimate students' attitudes and motivational beliefs (Dicke, Lüdtke, Trautwein, Nagy, & Nagy, 2012; Freiberger, Steinmayr, & Spinath, 2012; Machts, Kaiser, Schmidt, & Möller, 2016; Mullola, et al., 2014; Praetorius, Berner, Zeinz, Scheunpflug, & Dresel, 2013; Urhahne, Chao, Florineth, Luttenberger, & Paechter, 2011). For example, some teachers may have lower expectations or perceive lower abilities in students with different backgrounds, regardless of their attainment (Campbell, 2015; Strand, 2007); educators' and scientists' perceived images of ideal students or peers may perhaps limit their recognition of those with different genders or ethnicities (Carlone & Johnson, 2007). It may be broadly beneficial for teachers to self-reflect on their perceptions regarding their students, so that they can avoid biases and provide appropriate advice and guidance.

Similarly, it may be beneficial to help students' self-reflect and consider their own confidence and attainment. In broad terms, the results from the various accuracy/bias and intention/attitude clusters highlighted that few students exhibited high attainment and high accuracy; for the other students, however, accuracy/bias varied, suggesting that greater self-reflection may be beneficial. However, it remains somewhat unclear how this could be practically accomplished. This may be already undertaken to some extent in schools or other contexts, for example through students undertaking practice examination questions. Various studies have highlighted that interventions to promote confidence accuracy and/or self-regulated learning (involving some self-reflection) are broadly possible and have sometimes improved students' confidence accuracy, although results have varied (Bol & Hacker, 2001; Bol, Hacker, Walck, & Nunnery, 2012; Dignath & Büttner, 2008; Huff & Nietfeld, 2009; Labuhn, Zimmerman, & Hasselhorn, 2010). For example, such interventions have entailed benefits such as increased attainment (Bol, Hacker, Walck, & Nunnery, 2012; Kramarski, Mevarech, & Arami, 2002), although others have highlighted no effects (Bol, Hacker, O'Shea, & Allen, 2005; Gigerenzer, Hoffrage, & Kleinbölting, 1991). In broad terms, suggesting reflection on confidence and attainment does not necessarily entail that students must follow self-regulated learning or other formalised approaches, since the actual associations between accurate beliefs and self-regulated learning still remain somewhat unclear (Boekaerts, 1999; Pintrich, 1999; Stone, 2000).

Within science education, it perhaps remains somewhat unclear which attitudes, teaching practices, enrichment activities, and/or other related aspects should be promoted and how and why. Research studies have broadly associated students' science intentions with their attitudes such as their interest in science and perceived utility of science, together with their confidence and attainment (Bøe & Henriksen, 2015; Regan & DeWitt, 2015), as broadly seen in the results across this thesis. While students' intentions and aspirations have been found to be somewhat difficult to change directly, various interventions have instead shown the feasibility of increasing students' interest in science and perceived utility of science (Archer, DeWitt, & Dillon, 2014; Bernacki, Nokes-Malach, Richey, & Belenky, 2016; Harackiewicz, Rozek, Hulleman, & Hyde, 2012; Hong & Lin-Siegler, 2012; Hulleman & Harackiewicz, 2009; Rosenzweig & Wigfield, 2016; Rozek, Hyde, Svoboda, Hulleman, & Harackiewicz, 2015).

However, the identification of distinct clusters of students, each holding relatively similar within-cluster magnitudes of intentions and attitudes, suggested that students' attitudes may be more closely associated than previously considered. Accordingly, greater understanding of how students' attitudes associate and are influenced could then direct future interventions, and it is possible that interventions may need to address multiple factors concurrently. Additionally, and to reiterate, researchers and policy-makers may need to consider further how different students may make their choices in different ways; otherwise, any interventions to promote higher aspirations towards science may not necessarily achieve the desired effects.

Section 10.4: Conclusions

For Year 11 students in England surveyed in PISA 2006, science intentions were most strongly predicted by the students' perceived utility of science, interest/enjoyment of science, personal value of science, and their science self-concept beliefs (their subjective confidence in their abilities, measured through agreement or disagreement with items such as 'Science topics are easy for me' and 'I learn science topics quickly'). For students in Years 9, 10, and 11 surveyed in England in 2014/2015, science intentions were most strongly predicted by the students' perceived utility of science, personal value of science, self-efficacy (measured through confidence in their expected grades at GCSE/A-Level), subjective-norms/influences from parents, and the students' interest in science. These results reaffirmed the relevance of the perceived utility of science and students' confidence, but highlighted the importance of contextualisation in expressions of confidence, and also showed the relevance of considering the personal value of science to someone's identity through quantitative perspectives.

Students with different degrees of accuracy/bias expressed different science intentions and attitudes, and could be inferred to consider their choices in different ways. Emergent clusters of students could also be broadly described as showing tendencies towards under-confidence, accurately-low confidence, accurately-high confidence, or over-confidence. This affirmed the practice of classifying students into such conceptual groups, but the results highlighted the benefit of applying multiple approaches: the profiles and sizes of the conceptual groups and emergent clusters varied, and could entail wider differences in results.

Students with accurately-high confidence generally reported the highest intentions and attitudes, across both surveys and when considering conceptual groups or emergent clusters of students. In PISA 2006, under-confident students exhibited lower intentions and attitudes than over-confident students across the groups and clusters. In the 2014/2015 survey, considering task-level confidence accuracy/bias, results varied: the under-confident and over-confident conceptual groups exhibited similar science intentions and self-efficacy, but under-confident students exhibited lower attitudes including interest, utility, and personal value; conversely, the clusters showed that those exhibiting over-confidence reported lower intentions and attitudes than those exhibiting under-confidence. Groups/clusters with 'accurately-high' confidence could also or conversely appear over-confident, however, perhaps suggesting the contextual or conditional nature of any benefits or detriments of confidence biases, such as requiring some level of underlying attainment and/or confidence.

Concurrently, across both surveys, it can be inferred that the students' perceived utility value of science was more predictive of intentions for students with accurately-high confidence compared to other groups and clusters. Depending on whether or which groups and/or clusters were considered, utility value could be more predictive of intentions for under-confident students than over-confident students, and personal value could be more predictive for over-confident students than under-confident students, or that no differences could be observed.

While the differences across methods and surveys highlighted that further research would be needed to clarify such areas, these results provided a beneficial foundation, given little prior research into confidence accuracy/bias in England. Fundamentally, the results highlighted that the impact of under-confidence and over-confidence may be more complex than previously assumed, not simply through associating with lower or higher attitudes, but also through students considering their choices in different ways.

References

Abrahams, I. (2009). Does Practical Work Really Motivate? A study of the affective value of practical work in secondary school science. *International Journal of Science Education, 31*(17), 2335-2353. doi:10.1080/09500690802342836

Abrahams, I., & Reiss, M. J. (2012). Practical Work: Its Effectiveness in Primary and Secondary Schools in England. *Journal of Research in Science Teaching, 49*(8), 1035-1055. doi:10.1002/tea.21036

Abrahams, I., Reiss, M. J., & Sharpe, R. (2014). The impact of the 'Getting Practical: Improving Practical Work in Science' continuing professional development programme on teachers' ideas and practice in science practical work. *Research in Science and Technological Education, 32*(3), 263-280. doi:10.1080/02635143.2014.931841

Ackerman, P., & Wolman, S. (2007). Determinants and Validity of Self-Estimates of Abilities and Self-Concept Measures. *Journal of Experimental Psychology: Applied, 13*(2), 57-78. doi:10.1037/1076-898X.13.2.57

Ackerman, P., Beier, M., & Bowen, K. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences, 33*(4), 587-605. doi:10.1016/S0191-8869(01)00174-X

Aeschlimann, B., Herzog, W., & Makarova, E. (2016). How to foster students' motivation in mathematics and science classes and promote students' STEM career choice. A study in Swiss high schools. *International Journal of Educational Research, 79*, 31-41. doi:10.1016/j.ijer.2016.06.004

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211. doi:10.1016/0749-5978(91)90020-T

Ajzen, I. (2002). Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior. *Journal of Applied Social Psychology, 32*(4), 665-683. doi:10.1111/j.1559-1816.2002.tb00236.x

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723. doi:10.1109/TAC.1974.1100705

Alicke, M., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20*(1), 1-48. doi:10.1080/10463280802613866

Alicke, M., Klotz, M. L., Breitenbecher, D., Yurak, T., & Vredenburg, D. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology, 68*(5), 804-825. doi:10.1037/0022-3514.68.5.804

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 261-271. doi:10.1037/0022-0663.84.3.261

Anderhag, P., Wickman, P.-O., Bergqvist, K., Jakobson, B., Hamza, K. M., & Säljö, R. (2016). Why Do Secondary School Students Lose Their Interest in Science? Or Does it Never Emerge? A Possible and Overlooked Explanation. *Science Education, 100*(5), 791-813. doi:10.1002/sce.21231

Andersen, H. M., Krogh, L., & Lykkegaard, E. (2014). Identity Matching to Scientists: Differences that Make a Difference? *Research in Science Education, 44*(3), 439-460. doi:10.1007/s11165-013-9391-9

Andersen, L., & Chen, J. (2016). Do High-Ability Students Disidentify With Science? A Descriptive Study of U.S. Ninth Graders in 2009. *Science Education, 100*(1), 57-77. doi:10.1002/sce.21197

Andersen, L., & Cross, T. (2014). Are Students With High Ability in Math More Motivated in Math and Science Than Other Students? *Roeper Review, 36*(4), 221-234. doi:10.1080/02783193.2014.945221

Anderson, C., Brion, S., Moore, D., & Kennedy, J. (2012). A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology, 103*(4), 718-735. doi:10.1037/a0029395

Archambault, I., Eccles, J., & Vida, M. (2010). Ability Self-Concepts and Subjective Value in Literacy: Joint Trajectories From Grades 1 Through 12. *Journal of Educational Psychology, 102*(4), 804-816. doi:10.1037/a0021075

Archer, L., & DeWitt, J. (2015). Science Aspirations and Gender Identity: Lessons from the ASPIRES Project. In E. K. Henriksen, J. Dillon, & J. Ryder (Eds.), *Understanding Student Participation and Choice in Science and Technology Education* (pp. 89-102). Dordrecht: Springer. doi:10.1007/978-94-007-7793-4_6

Archer, L., & Francis, B. (2006). Challenging Classes? Exploring the Role of Social Class within the Identities and Achievement of British Chinese Pupils. *Sociology, 40*(1), 29-49. doi:10.1177/0038038506058434

Archer, L., Dawson, E., DeWitt, J., Seakins, A., & Wong, B. (2015). "Science capital": A conceptual, methodological, and empirical argument for extending Bourdieusian notions of capital beyond the arts. *Journal of Research in Science Teaching, 52*(7), 922-948. doi:10.1002/tea.21227

Archer, L., DeWitt, J., & Dillon, J. (2014). 'It didn't really change my opinion': exploring what works, what doesn't and why in a school science, technology, engineering and mathematics careers intervention. *Research in Science & Technological Education, 32*(1), 35-55. doi:10.1080/02635143.2013.865601

Archer, L., DeWitt, J., & Osborne, J. (2015). Is Science for Us? Black Students' and Parents' Views of Science and Science Careers. *Science Education, 99*(2), 199-237. doi:10.1002/sce.21146

Archer, L., DeWitt, J., & Wong, B. (2014). Spheres of influence: what shapes young people's aspirations at age 12/13 and what are the implications for education policy? *Journal of Education Policy, 29*(1), 58-85. doi:10.1080/02680939.2013.790079

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2010). "Doing" science versus "being" a scientist: Examining 10/11-year-old schoolchildren's constructions of science through the lens of identity. *Science Education, 94*(4), 617-639. doi:10.1002/sce.20399

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2012). Science Aspirations, Capital, and Family Habitus: How Families Shape Children's Engagement and Identification With Science. *American Educational Research Journal, 49*(5), 881-908. doi:10.3102/0002831211433290

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2013). 'Not girly, not sexy, not glamorous': primary school girls' and parents' constructions of science aspirations. Pedagogy, *Culture & Society, 21*(1), 171-194. doi:10.1080/14681366.2012.748676

Archer, L., Halsall, A., & Hollingworth, S. (2007). Class, gender, (hetero)sexuality and schooling: paradoxes within working-class girls' engagement with education and post-16 aspirations. *British Journal of Sociology of Education, 28*(2), 165-180. doi:10.1080/01425690701192570

Arens, A. K., Seeshing Yeung, A., Craven, R., & Hasselhorn, M. (2011). The Twofold Multidimensionality of Academic Self-Concept: Domain Specificity and Separation Between Competence and Affect Components. *Journal of Educational Psychology, 103*(4), 970-981. doi:10.1037/a0025047

Arens, A. K., Seeshing Yeung, A., Craven, R., & Watermann, R. (2013). Does the timing of transition matter? Comparison of German students' self-perceptions before and after transition to secondary school. *International Journal of Educational Research, 57*, 1-11. doi:10.1016/j.ijer.2012.11.001

Armstrong, N., Wallace, C., & Chang, S.-M. (2008). Learning from Writing in College Biology. *Research in Science Education, 38*(4), 483-499. doi:10.1007/s11165-007-9062-9

Aschbacher, P., Ing, M., & Tsai, S. (2014). Is Science Me? Exploring Middle School Students' STE-M Career Aspirations. *Journal of Science Education and Technology, 23*(6), 735-743. doi:10.1007/s10956-014-9504-x

Aschbacher, P., Li, E., & Roth, E. (2010). Is Science Me? High School Students' Identities, Participation and Aspirations in Science, Engineering, and Medicine. *Journal of Research in Science Teaching, 47*(5), 564-582. doi:10.1002/tea.20353

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64*(6), 359-372. doi:10.1037/h0043445

Atkinson, J. W. (1964). *An Introduction to Motivation*. Princeton, New Jersey: Van Nostrand.

Attwood, G., Croll, P., Fuller, C., & Last, K. (2013). The accuracy of students' predictions of their GCSE grades. *Educational Studies, 39*(4), 444-454. doi:10.1080/03055698.2013.776945

Audi, R. (1998). *Epistemology: A contemporary introduction to the theory of knowledge*. London: Routledge.

Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review, 84*(2), 191-215. doi:10.1037/0033-295X.84.2.191

Bandura, A. (1986). *Social foundations of thought and action: a social cognitive theory*. Englewood Cliffs, New Jersey: Prentice-Hall.
Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44*(9), 1175-1184. doi:10.1037/0003-066X.44.9.1175
Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
Bandura, A. (2001). Social Cognitive Theory: An Agentic Perspective. *Annual Review of Psychology, 52*, 1-26. doi:10.1146/annurev.psych.52.1.1
Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-Efficacy Beliefs as Shapers of Children's Aspirations and Career Trajectories. *Child Development, 72*(1), 187-206. doi:10.1111/1467-8624.00273
Barnette, J. J. (2000). Effects of Stem and Likert Response Option Reversals on Survey Internal Consistency: If You Feel the Need, There is a Better Alternative to Using those Negatively Worded Stems. *Educational and Psychological Measurement, 60*(3), 361-370. doi:10.1177/00131640021970592
Baron, R., & Kenny, D. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182. doi:10.1037/0022-3514.51.6.1173
Bartholomew, D., Steele, F., Moustaki, I., & Galbraith, J. (2008). *Analysis of Multivariate Social Science Data* (2nd ed.). Boca Raton, Florida: Chapman & Hall/CRC.
Bates, P., Pollard, E., Usher, T., & Oakley, J. (2009). *Who is Heading for HE? Young People's Perceptions of, and Decisions About, Higher Education*. London: Department for Business, Innovation and Skills.
Bennett, J., & Hogarth, S. (2009). Would You Want to Talk to a Scientist at a Party? High school students' attitudes to school science and to science. *International Journal of Science Education, 31*(14), 1975-1998. doi:10.1080/09500690802425581
Bennett, J., Lubben, F., & Hampden-Thompson, G. (2013). Schools That Make a Difference to Post-Compulsory Uptake of Physical Science Subjects: Some comparative case studies in England. *International Journal of Science Education, 35*(4), 663-689. doi:10.1080/09500693.2011.641131
Benson, J., & Hocevar, D. (1985). The Impact of Item Phrasing on the Validity of Attitude Scales for Elementary School Children. *Journal of Educational Measurement, 22*(3), 231-240. doi:10.1111/j.1745-3984.1985.tb01061.x
Bernacki, M., Nokes-Malach, T., Richey, J. E., & Belenky, D. (2016). Science diaries: a brief writing intervention to improve motivation to learn science. *Educational Psychology, 36*(1), 26-46. doi:10.1080/01443410.2014.895293
Biggs, J. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology, 63*(1), 3-19. doi:10.1111/j.2044-8279.1993.tb01038.x
Bipp, T., Steinmayr, R., & Spinath, B. (2012). A functional look at goal orientations: Their role for self-estimates of intelligence and performance. *Learning and Individual Differences, 22*(3), 280-289. doi:10.1016/j.lindif.2012.01.009
Blatchford, P. (1997). Pupils' self assessments of academic attainment at 7, 11 and 16 years: effects of sex and ethnic group. *British Journal of Educational Psychology, 67*(2), 169-184. doi:10.1111/j.2044-8279.1997.tb01235.x
Blenkinsop, S., McCrone, T., Wade, P., & Morris, M. (2006). *How do young people make choices at 14 and 16? DfES Research Report 773*. Nottingham: Department for Education and Skills.
Blickenstaff, J. C. (2005). Women and science careers: leaky pipeline or gender filter? *Gender and Education, 17*(4), 369-386. doi:10.1080/09540250500145072
Bøe, M. V. (2012). Science Choices in Norwegian Upper Secondary School: What Matters? *Science Education, 96*(1), 1-20. doi:10.1002/sce.20461
Bøe, M. V., & Henriksen, E. K. (2013). Love It or Leave It: Norwegian Students' Motivations and Expectations for Postcompulsory Physics. *Science Education, 97*(4), 550-573. doi:10.1002/sce.21068
Bøe, M. V., & Henriksen, E. K. (2015). Expectancy-Value Perspectives on Choice of Science and Technology Education in Late-Modern Societies. In E. K. Henriksen, J. Dillon, & J. Ryder (Eds.), *Understanding Student Participation and Choice in Science and Technology Education* (pp. 17-29). Dordrecht: Springer. doi:10.1007/978-94-007-7793-4_2
Bøe, M. V., Henriksen, E. K., Lyons, T., & Schreiner, C. (2011). Participation in science and technology: young people's achievement-related choices in late-modern

societies. *Studies in Science Education, 47*(1), 37-72. doi:10.1080/03057267.2011.549621
Boekaerts, M. (1999). Self-regulated learning: where we are today. *International Journal of Educational Research, 31*(6), 445-457. doi:10.1016/S0883-0355(99)00014-2
Boekaerts, M., & Rozendaal, J. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction, 20*(5), 372-382. doi:10.1016/j.learninstruc.2009.03.002
Bol, L., & Hacker, D. (2001). A Comparison of the Effects of Practice Tests and Traditional Review on Performance and Calibration. *The Journal of Experimental Education, 69*(2), 133-151. doi:10.1080/00220970109600653
Bol, L., Hacker, D., O'Shea, P., & Allen, D. (2005). The Influence of Overt Practice, Achievement Level, and Explanatory Style on Calibration Accuracy and Performance. *The Journal of Experimental Education, 73*(4), 269-290. doi:10.3200/JEXE.73.4.269-290
Bol, L., Hacker, D., Walck, C., & Nunnery, J. (2012). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemporary Educational Psychology, 37*(4), 280-287. doi:10.1016/j.cedpsych.2012.02.004
Bollen, K. (2002). Latent Variables in Psychology and the Social Sciences. *Annual Review of Psychology, 53*, 605-634. doi:10.1146/annurev.psych.53.100901.135239
Bong, M. (1996). Problems in Academic Motivation Research and Advantages and Disadvantages of Their Solutions. *Contemporary Educational Psychology, 21*(2), 149-165. doi:10.1006/ceps.1996.0013
Bong, M. (1997). Generality of Academic Self-Efficacy Judgments: Evidence of Hierarchical Relations. *Journal of Educational Psychology, 89*(4), 696-709. doi:10.1037/0022-0663.89.4.696
Bong, M. (1998). Tests of the Internal/External Frames of Reference Model With Subject-Specific Academic Self-Efficacy and Frame-Specific Academic Self-Concepts. *Journal of Educational Psychology, 90*(1), 102-110. doi:10.1037/0022-0663.90.1.102
Bong, M. (2001a). Between- and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task value, and achievement goals. *Journal of Educational Psychology, 93*(1), 23-34. doi:10.1037/0022-0663.93.1.23
Bong, M. (2001b). Role of Self-Efficacy and Task-Value in Predicting College Students' Course Performance and Future Enrollment Intentions. *Contemporary Educational Psychology*, *26*(4), 553-570. doi:10.1006/ceps.2000.1048
Bong, M. (2002). Predictive Utility of Subject-, Task-, and Problem-Specific Self-Efficacy Judgments for Immediate and Delayed Academic Performances. *The Journal of Experimental Education, 70*(2), 133-162. doi:10.1080/00220970209599503
Bong, M. (2004). Academic Motivation in Self-Efficacy, Task Value, Achievement Goal Orientations, and Attributional Beliefs. *The Journal of Educational Research, 97*(6), 287-298. doi:10.3200/JOER.97.6.287-298
Bong, M., & Clark, R. (1999). Comparison Between Self-concept and Self-Efficacy in Academic Motivation Research. *Educational Psychologist, 34*(3), 139-153. doi:10.1207/s15326985ep3403_1
Bong, M., & Skaalvik, E. (2003). Academic Self-Concept and Self-Efficacy: How Different Are They Really? *Educational Psychology Review, 15*(1), 1-40. doi:10.1023/A:1021302408382
Borsboom, D., Mellenbergh, G., & van Heerden, J. (2003). The Theoretical Status of Latent Variables. *Psychological Review, 110*(2), 203-219. doi:10.1037/0033-295X.110.2.203
Bosker, R., & Dekkers, H. (1994). School Differences in Producing Gender-Related Subject Choices. *School Effectiveness and School Improvement, 5*(2), 178-195. doi:10.1080/0924345940050204
Bosworth, D., Lyonette, C., Wilson, R., Bayliss, M., & Fathers, S. (2013). *The Supply of and Demand for High-Level STEM Skills*. London: UK Commission for Employment and Skills.
Bouffard, T., & Narciss, S. (2011). Benefits and risks of positive biases in self-evaluation of academic competence: Introduction. *International Journal of Educational Research, 50*(4), 205-208. doi:10.1016/j.ijer.2011.08.001

Bouffard, T., Boisvert, M., & Vezeau, C. (2003). The illusion of incompetence and its correlates among elementary school children and their parents. *Learning and Individual Differences, 14*(1), 31-46. doi:10.1016/j.lindif.2003.07.001

Bouffard, T., Vezeau, C., Roy, M., & Lengelé, A. (2011). Stability of biases in self-evaluation and relations to well-being among elementary school children. *International Journal of Educational Research, 50*(4), 221-229. doi:10.1016/j.ijer.2011.08.003

Bourdieu, P. (1984). *Distinction: a social critique of the judgement of taste*. (R. Nice, Trans.) Cambridge, Massachusetts: Harvard University Press.

Bradshaw, J., Sturman, L., Vappula, H., Ager, R., & Wheater, R. (2007). *Achievement of 15-year-olds in England: PISA 2006 National Report*. Slough: National Foundation for Educational Research.

Brandell, G., & Staberg, E.-M. (2008). Mathematics: a female, male or gender-neutral domain? A study of attitudes among students at secondary level. *Gender and Education, 20*(5), 495-509. doi:10.1080/09540250701805771

Brenner, L. (2000). Should observed overconfidence be dismissed as a statistical artifact? Critique of Erev, Wallsten, and Budescu (1994). *Psychological Review, 107*(4), 943-946. doi:10.1037/0033-295X.107.4.943

Brier, G. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review, 78*(1), 1-3. doi:10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

British Academy. (2015). *Count us in: Quantitative skills for a new generation*. London: British Academy.

Britner, S. (2008). Motivation in High School Science Students: A Comparison of Gender Differences in Life, Physical, and Earth Science Classes. *Journal of Research in Science Teaching, 45*(8), 955-970. doi:10.1002/tea.20249

Britner, S., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching, 43*(5), 485-499. doi:10.1002/tea.20131

Brown, M., Brown, P., & Bibby, T. (2008). “I would rather die”: reasons given by 16-year-olds for not continuing their study of mathematics. *Research in Mathematics Education, 10*(1), 3-18. doi:10.1080/14794800801915814

Budescu, D., Wallsten, T., & Au, W. T. (1997). On the Importance of Random Error in the Study of Probability Judgment. Part II: Applying the Stochastic Judgment Model to Detect Systematic Trends. *Journal of Behavioral Decision Making, 10*(3), 173-188. doi:10.1002/(SICI)1099-0771(199709)10:3<173::AID-BDM261>3.0.CO;2-6

Burnette, J., O’Boyle, E., VanEpps, E., Pollack, J., & Finkel, E. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin, 139*(3), 655-701. doi:10.1037/a0029531

Buschor, C. B., Berweger, S., Frei, A. K., & Kappler, C. (2014). Majoring in STEM—What Accounts for Women’s Career Decision Making? A Mixed Methods Study. *The Journal of Educational Research, 107*(3), 167-176. doi:10.1080/00220671.2013.788989

Butler, D., & Winne, P. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research, 65*(3), 245-281. doi:10.3102/00346543065003245

Butz, A., & Usher, E. (2015). Salient sources of early adolescents’ self-efficacy in two domains. *Contemporary Educational Psychology, 42*, 49-61. doi:10.1016/j.cedpsych.2015.04.001

Byars-Winston, A., Estrada, Y., Howard, C., Davis, D., & Zalapa, J. (2010). Influence of Social Cognitive and Ethnic Variables on Academic Goals of Underrepresented Students in Science and Engineering: A Multiple-Groups Analysis. *Journal of Counseling Psychology, 57*(2), 205-218. doi:10.1037/a0018608

Cai, H., Sedikides, C., Gaertner, L., Wang, C., Carvallo, M., Xu, Y., . . . Eckstein Jackson, L. (2011). Tactical Self-Enhancement in China: Is Modesty at the Service of Self-Enhancement in East Asian Culture? *Social Psychological and Personality Science, 2*(1), 59-64. doi:10.1177/1948550610376599

Calabrese Barton, A., Kang, H., Tan, E., O’Neill, T., Bautista-Guerra, J., & Brecklin, C. (2013). Crafting a Future in Science: Tracing Middle School Girls’ Identity Work Over Time and Space. *American Educational Research Journal, 50*(1), 37-75. doi:10.3102/0002831212458142

Campbell, T. (2015). Stereotyped at Seven? Biases in Teacher Judgement of Pupils' Ability and Attainment. *Journal of Social Policy, 44*(3), 517-547. doi:10.1017/S0047279415000227
Cann, R. (2009). Girls' participation in post-16 mathematics: a view from pupils in Wales. *Gender in Education, 21*(6), 651-669. doi:10.1080/09540250802680065
Cannon, W. (1929). Organization for physiological homeostasis. *Physiological Reviews, 9*(3), 399-431.
Carlone, H., & Johnson, A. (2007). Understanding the Science Experiences of Successful Women of Color: Science Identity as an Analytic Lens. *Journal of Research in Science Teaching, 44*(8), 1187-1218. doi:10.1002/tea.20237
Carlone, H., Johnson, A., & Scott, C. (2015). Agency Amidst Formidable Structures: How Girls Perform Gender in Science Class. *Journal of Research in Science Teaching, 52*(4), 474-488. doi:10.1002/tea.21224
Carlone, H., Webb, A., Archer, L., & Taylor, M. (2015). What Kind of Boy Does Science? A Critical Perspective on the Science Trajectories of Four Scientifically Talented Boys. *Science Education, 99*(3), 438-464. doi:10.1002/sce.21155
Carroll, J. (1989). The Carroll Model: A 25-Year Retrospective and Prospective View. *Educational Researcher, 18*(1), 26-31. doi:10.3102/0013189X018001026
Carver, C., & Scheier, M. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin, 92*(1), 111-135. doi:10.1037/0033-2909.92.1.111
CASE [Campaign for Science and Engineering]. (2014). *Improving Diversity in STEM*. London: Campaign for Science and Engineering.
Cavagnetto, A. (2010). Argument to Foster Scientific Literacy: A Review of Argument Interventions in K–12 Science Contexts. *Review of Educational Research, 80*(3), 336-371. doi:10.3102/0034654310376953
Chatard, A., Guimond, S., & Selimbegovic, L. (2007). "How good are you in math?" The effect of gender stereotypes on students' recollection of their school marks. *Journal of Experimental Social Psychology, 43*(6), 1017-1024. doi:10.1016/j.jesp.2006.10.024
Cheema, J., & Skultety, L. (2016). Self-efficacy and literacy: a paired difference approach to estimation of over-/under-confidence in mathematics- and science-related tasks. *Educational Psychology*, 1-14. doi:10.1080/01443410.2015.1127329
Chen, J., & Pajares, F. (2010). Implicit theories of ability of Grade 6 science students: Relation to epistemological beliefs and academic motivation and achievement in science. *Contemporary Educational Psychology, 35*(1), 75-87. doi:10.1016/j.cedpsych.2009.10.003
Chen, J., & Usher, E. (2013). Profiles of the sources of science self-efficacy. *Learning and Individual Differences, 24*, 11-21. doi:10.1016/j.lindif.2012.11.002
Chen, P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences, 14*(1), 79-92. doi:10.1016/j.lindif.2003.08.003
Chen, P., & Zimmerman, B. (2007). A Cross-National Comparison Study on the Accuracy of Self-Efficacy Beliefs of Middle-School Mathematics Students. *The Journal of Experimental Education, 75*(3), 221-244. doi:10.3200/JEXE.75.3.221-244
Chen, S.-K., Yeh, Y.-C., Hwang, F.-M., & Lin, S. (2013). The relationship between academic self-concept and achievement: A multicohort–multioccasion study. *Learning and Individual Differences, 23*, 172-178. doi:10.1016/j.lindif.2012.07.021
Chevalier, A., Gibbons, S., Thorpe, A., Snelle, M., & Hoskins, S. (2009). Students' academic self-perception. *Economics of Education Review, 28*(6), 716-727. doi:10.1016/j.econedurev.2009.06.007
Chiu, D., Beru, Y., Watley, E., Wubu, S., Simson, E., Kessinger, R., . . . Wigfield, A. (2008). Influences of Math Tracking on Seventh-Grade Students' Self-Beliefs and Social Comparisons. *The Journal of Educational Research, 102*(2), 125-136. doi:10.3200/JOER.102.2.125-136
Chiu, M. M., & Klassen, R. (2009). Calibration of reading self-concept and reading achievement among 15-year-olds: Cultural differences in 34 countries. *Learning and Individual Differences, 19*(3), 372-386. doi:10.1016/j.lindif.2008.10.004
Chiu, M. M., & Klassen, R. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction, 20*(1), 2-17. doi:10.1016/j.learninstruc.2008.11.002

Chiu, M. M., Chow, B. W.-Y., & Mcbride-Chang, C. (2007). Universals and specifics in learning strategies: Explaining adolescent mathematics, science, and reading achievement across 34 countries. *Learning and Individual Differences, 17*(4), 344-365. doi:10.1016/j.lindif.2007.03.007
Chiu, M.-S. (2008). Achievements and self-concepts in a comparison of math and science: exploring the internal/external frame of reference model across 28 countries. *Educational Research and Evaluation, 14*(3), 235-254. doi:10.1080/13803610802048858
Chiu, M.-S. (2012). Differential Psychological Processes Underlying the Skill-Development Model and Self-Enhancement Model across Mathematics and Science in 28 Countries. *International Journal of Science and Mathematics Education, 10*(3), 611-642. doi:10.1007/s10763-011-9309-9
Chow, A., Eccles, J., & Salmela-Aro, K. (2012). Task Value Profiles Across Subjects and Aspirations to Physical and IT-Related Sciences in the United States and Finland. *Developmental Psychology, 48*(6), 1612-1628. doi:10.1037/a0030194
Church, A. T., Katigbak, M., Mazuera Arias, R., Rincon, B. C., Vargas-Flores, J. D., Ibáñez-Reyes, J., . . . Ortiz, F. (2014). A four-culture study of self-enhancement and adjustment using the social relations model: Do alternative conceptualizations and indices make a difference? *Journal of Personality and Social Psychology, 106*(6), 997-1014. doi:10.1037/a0036491
Clark, L. A., & Watson, D. (1995). Constructing Validity: Basic Issues in Objective Scale Development. *Psychological Assessment, 7*(3), 309-319. doi:10.1037/1040-3590.7.3.309
Claussen, S., & Osborne, J. (2013). Bourdieu's notion of cultural capital and its implications for the science curriculum. *Science Education, 97*(1), 58-79. doi:10.1002/sce.21040
Cleary, T., & Chen, P. (2009). Self-regulation, motivation, and math achievement in middle school: Variations across grade level and math context. *Journal of School Psychology, 47*(5), 291-314. doi:10.1016/j.jsp.2009.04.002
Cleaves, A. (2005). The formation of science choices in secondary school. *International Journal of Science Education, 27*(4), 471-486. doi:10.1080/0950069042000323746
Coe, R. (2008). Comparability of GCSE examinations in different subjects: an application of the Rasch model. *Oxford Review of Education, 34*(5), 609-636. doi:10.1080/03054980801970312
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Earlbaum Associates.
Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159. doi:10.1037/0033-2909.112.1.155
Cohen, J., & Cohen, P. (1984). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education* (6th ed.). Oxford: Routledge.
Collins, L., & Lanza, S. (2010). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. Hoboken, New Jersey: Wiley.
Collins, R. (1996). For better or worse: The impact of upward social comparison on self-evaluations. *Psychological Bulletin, 119*(1), 51-69. doi:10.1037/0033-2909.119.1.51
Conley, A. (2012). Patterns of Motivation Beliefs: Combining Achievement Goal and Expectancy-Value Perspectives. *Journal of Educational Psychology, 104*(1), 32-47. doi:10.1037/a0026042
Connor, H., Tyers, C., Modood, T., & Hillage, J. (2004). *Why the Difference? A Closer Look at Higher Education Minority Ethnic Students and Graduates*. *Research Report RR552*. Nottingham: DfES Publications.
Correll, S. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology, 106*(6), 1691-1730. doi:10.1086/321299
Côté, J. (2006). Identity Studies: How Close Are We to Developing a Social Science of Identity? - An Appraisal of the Field. *Identity, 6*(1), 3-25. doi:10.1207/s1532706xid0601_2

Crawford, C. (2014). *The link between secondary school characteristics and university participation and outcomes. DFE-RR353*. London: Department for Education.
Credé, M., & Phillips, L. A. (2011). A meta-analytic review of the Motivated Strategies for Learning Questionnaire. *Learning and Individual Differences, 21*(4), 337-346. doi:10.1016/j.lindif.2011.03.002
Crombie, G., Sinclair, N., Silverthorn, N., Byrne, B., DuBois, D., & Trinneer, A. (2005). Predictors of Young Adolescents' Math Grades and Course Enrollment Intentions: Gender Similarities and Differences. *Sex Roles, 52*(5/6), 351-367. doi:10.1007/s11199-005-2678-1
Cronbach, L. (1950). Further Evidence on Response Sets and Test Design. *Educational and Psychological Measurement, 10*(1), 3-31. doi:10.1177/001316445001000101
Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. doi:10.1007/BF02310555
Cronbach, L., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin, 74*(1), 68-80. doi:10.1037/h0029382
Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302. doi:10.1037/h0040957
Cundiff, J., Vescio, T., Loken, E., & Lo, L. (2013). Do gender–science stereotypes predict science identification and science career aspirations among undergraduate science majors? *Social Psychology of Education, 16*(4), 541-554. doi:10.1007/s11218-013-9232-8
Dabney, K., Chakraverty, D., & Robert, T. (2013). The Association of Family Influence and Initial Interest in Science. *Science Education, 97*(3), 395-409. doi:10.1002/sce.21060
Davies, P., Qiu, T., & Davies, N. (2014). Cultural and human capital, information and higher education choices. *Journal of Education Policy, 29*(6), 804-825. doi:10.1080/02680939.2014.891762
de Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York: The Guilford Press.
Deci, E., & Ryan, R. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.
Dempster, A., Laird, N., & Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological), 39*(1), 1-38.
Denissen, J., Zarrett, N., & Eccles, J. (2007). I Like to Do It, I'm Able, and I Know I Am: Longitudinal Couplings Between Domain-Specific Achievement, Self-Concept, and Interest. *Child Development, 78*(2), 430-447. doi:10.1111/j.1467-8624.2007.01007.x
Department for Children, Schools and Families. (2009). *Progression to post-16 science: the report*. London: Department for Children, Schools and Families.
Department for Education. (2011). *Maths and science education: the supply of high achievers at A level. Research Report DFE-RR079*. London: Department for Education.
Department for Education. (2012). *Subject progression from GCSE to AS Level and continuation to A Level. Research Report DFE-RR195*. London: Department for Education.
Department for Education. (2013). *National curriculum in England: science programmes of study*. London: Department for Education. Retrieved July 1, 2016, from https://www.gov.uk/government/publications/national-curriculum-in-england-science-programmes-of-study
Department for Education. (2014). *The national curriculum in England: Key stages 3 and 4 framework document*. London: Department for Education. Retrieved July 1, 2016, from https://www.gov.uk/government/publications/national-curriculum-in-england-secondary-curriculum
Department for Education. (2015). *Careers guidance and inspiration in schools: Statutory guidance for governing bodies, school leaders and school staff*. London: Department for Education. Retrieved July 1, 2016, from https://www.gov.uk/government/publications/careers-guidance-provision-for-young-people-in-schools
Department of Education. (2016, July 1). *Compare school and college performance*. Retrieved August 31, 2016, from Department of Education: https://www.compare-school-performance.service.gov.uk/download-data

Destan, N., & Roebers, C. (2015). What are the metacognitive costs of young children's overconfidence? *Metacognition Learning, 10*(3), 347-374. doi:10.1007/s11409-014-9133-z

DeWitt, J., & Archer, L. (2015). Who Aspires to a Science Career? A comparison of survey responses from primary and secondary school students. *International Journal of Science Education, 37*(13), 2170-2192. doi:10.1080/09500693.2015.1071899

DeWitt, J., Archer, L., & Osborne, J. (2014). Science-related Aspirations Across the Primary–Secondary Divide: Evidence from two surveys in England. *International Journal of Science Education, 36*(10), 1609-1629. doi:10.1080/09500693.2013.871659

DeWitt, J., Archer, L., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2011). High aspirations but low progression: the science aspirations-careers paradox amongst minority ethnic students. *International Journal of Science and Mathematics Education, 9*(2), 243-271. doi:10.1007/s10763-010-9245-0

Dicke, A.-L., Lüdtke, O., Trautwein, U., Nagy, G., & Nagy, N. (2012). Judging students' achievement goal orientations: Are teacher ratings accurate? *Learning and Individual Differences, 22*(6), 844-849. doi:10.1016/j.lindif.2012.04.004

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3*(3), 231-264. doi:10.1007/s11409-008-9029-x

DiStefano, C., & Kamphaus, R. (2006). Investigating Subtypes of Child Development: A Comparison of Cluster Analysis and Latent Class Cluster Analysis in Typology Creation. *Educational and Psychological Measurement, 66*(5), 778-794. doi:10.1177/0013164405284033

Dupeyrat, C., Escribe, C., Huet, N., & Régner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *International Journal of Educational Research, 50*(4), 241-250. doi:10.1016/j.ijer.2011.08.005

Dweck, C. (2000). *Self-Theories: Their Role in Motivation, Personality, and Development*. Philadelphia: Psychology Press.

Dweck, C., & Leggett, E. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256-273. doi:10.1037/0033-295X.95.2.256

EACEA [Education, Audiovisual and Culture Executive Agency]. (2011). *Science Education in Europe: National Policies, Practices and Research*. Brussels: European Commission. doi:10.2797/7170

Eccles, J. (2009). Who Am I and What Am I Going to Do With My Life? Personal and Collective Identities as Motivators of Action. *Educational Psychologist, 44*(2), 78-89. doi:10.1080/00461520902832368

Eccles, J., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology, 53*, 109-132. doi:10.1146/annurev.psych.53.100901.135153

Eccles, J., Adler, T., Futterman, R., Goff, S., Kaczala, C., Meece, J., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. Spence (Ed.), *Achievement and Achievement Motives* (pp. 75-146). San Francisco: W. H. Freeman.

Echazarra, A., Salinas, D., Méndez, I., Denis, V., & Rech, G. (2016). *How teachers teach and learn: Successful strategies for school*. Paris: OECD Publishing. doi:10.1787/5jm29kpt0xxx-en

Edgerton, J., & Roberts, L. (2014). Cultural capital or habitus? Bourdieu and beyond in the explanation of enduring educational inequality. *Theory and Research in Education, 12*(2), 193-220. doi:10.1177/1477878514530231

Edwards, J., & Bagozzi, R. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*(2), 155-174. doi:10.1037/1082-989X.5.2.155

Efklides, A., Kourkoulou, A., Mitsiou, F., & Ziliaskopoulou, D. (2006). Metacognitive knowledge of effort, personality factors, and mood state: their relationships with effort-related metacognitive experiences. *Metacognition Learning, 1*(1), 33-49. doi:10.1007/s11409-006-6581-0

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105*(1), 98-121. doi:10.1016/j.obhdp.2007.05.002

Ehrlinger, J., Mitchum, A., & Dweck, C. (2016). Understanding overconfidence: Theories of intelligence, preferential attention, and distorted self-assessment. *Journal of Experimental Social Psychology, 63,* 94-100. doi:10.1016/j.jesp.2015.11.001
Elliot, A. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist, 34*(3), 169-189. doi:10.1207/s15326985ep3403_3
Elliot, A., & McGregor, H. (2001). A 2 × 2 Achievement Goal Framework. *Journal of Personality and Social Psychology, 80*(3), 501-519. doi:10.1037//0022-3514.80.3.501
Elliot, A., & Thrash, T. (2001). Achievement Goals and the Hierarchical Model of Achievement Motivation. *Educational Psychology Review, 13*(2), 139-156. doi:10.1023/A:1009057102306
Else-Quest, N., Hyde, J. S., & Linn, M. (2010). Cross-National Patterns of Gender Differences in Mathematics: A Meta-Analysis. *Psychological Bulletin, 136*(1), 103-127. doi:10.1037/a0018053
Engineering UK. (2015). *Engineering UK 2015: The state of engineering*. London: Engineering UK.
Entwistle, N., & McCune, V. (2004). The Conceptual Bases of Study Strategy Inventories. *Educational Psychology Review, 16*(4), 325-345. doi:10.1007/s10648-004-0003-0
Erev, I., Wallsten, T., & Budescu, D. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101*(3), 519-527. doi:10.1037/0033-295X.101.3.519
Erickson, S., & Heit, E. (2015). Metacognition and confidence: comparing math to other academic subjects. *Frontiers in Psychology, 6*(742), 1-10. doi:10.3389/fpsyg.2015.00742
Espinoza, P., Arêas da Luz Fontes, A., & Arms-Chavez, C. (2014). Attributional gender bias: teachers' ability and effort explanations for students' math performance. *Social Psychology of Education, 17*(1), 105-126. doi:10.1007/s11218-013-9226-6
Eurydice. (2016, March 1). *Countries: Descriptions of national educational systems*. Retrieved April 18, 2016, from Eurydice: https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Countries
Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th ed.). Chichester: Wiley.
Federici, R., & Skaalvik, E. (2014). Students' perception of instrumental support and effort in mathematics: the mediating role of subjective task values. *Social Psychology of Education, 17*(3), 527-540. doi:10.1007/s11218-014-9264-8
Ferguson, L., & Bråten, I. (2013). Student profiles of knowledge and epistemic beliefs: Changes and relations to multiple-text comprehension. *Learning and Instruction, 25*, 49-61. doi:10.1016/j.learninstruc.2012.11.003
Finn, B., & Metcalfe, J. (2007). The Role of Memory for Past Test in the Underconfidence With Practice Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 238-244. doi:10.1037/0278-7393.33.1.238
Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language, 58*(1), 19-34. doi:10.1016/j.jml.2007.03.006
Finn, B., & Metcalfe, J. (2014). Overconfidence in children's multi-trial judgments of learning. *Learning and Instruction, 32*, 1-9. doi:10.1016/j.learninstruc.2014.01.001
Fischbach, A., Keller, U., Preckel, F., & Brunner, M. (2013). PISA proficiency scores predict educational outcomes. *Learning and Individual Differences, 24*, 63-72. doi:10.1016/j.lindif.2012.10.012
Flake, J. K., Barron, K., Hulleman, C., McCoach, B., & Welsh, M. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology, 41*, 232-244. doi:10.1016/j.cedpsych.2015.03.002
Forgasz, H., Leder, G., & Kloosterman, P. (2004). New Perspectives on the Gender Stereotyping of Mathematics. *Mathematical Thinking and Learning, 6*(4), 389-420. doi:10.1207/s15327833mtl0604_2
Foy, P., Arora, A., & Stanco, G. (Eds.). (2013). *TIMSS 2011 User Guide for the International Database*. Chestnut Hill, Massachusetts: TIMSS & PIRLS International Study Centre.
Fredricks, J., & Eccles, J. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*(4), 519-533. doi:10.1037/0012-1649.38.4.519
Freiberger, V., Steinmayr, R., & Spinath, B. (2012). Competence beliefs and perceived ability evaluations: How do they contribute to intrinsic motivation and

achievement? *Learning and Individual Differences*, *22*(4), 518-522. doi:10.1016/j.lindif.2012.02.004
Frenzel, A., Pekrun, R., Dicke, A.-L., & Goetz, T. (2012). Beyond Quantitative Decline: Conceptual Shifts in Adolescents' Development of Interest in Mathematics. *Developmental Psychology, 48*(4), 1069-1082. doi:10.1037/a0026895
Freund, P. A., & Kasten, N. (2012). How Smart Do You Think You Are? A Meta-Analysis on the Validity of Self-Estimates of Cognitive Ability. *Psychological Bulletin, 138*(2), 296-321. doi:10.1037/a0026556
Gaertner, L., Sedikides, C., & Chang, K. (2008). On Pancultural Self-Enhancement: Well-Adjusted Taiwanese Self-Enhance on Personally Valued Traits. *Journal of Cross-Cultural Psychology, 39*(4), 463-477. doi:10.1177/0022022108318431
Ganzeboom, H., De Graaf, P., & Treiman, D. (1992). A standard international socio-economic index of occupational status. *Social Science Research, 21*(1), 1-56. doi:10.1016/0049-089X(92)90017-B
Gee, J. (2000). Identity as an Analytic Lens for Research in Education. *Review of Research in Education, 25*(1), 99-125. doi:10.3102/0091732X025001099
Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*(4), 506-528. doi:10.1037/0033-295X.98.4.506
Gill, T., & Bell, J. (2013). What Factors Determine the Uptake of A-level Physics? *International Journal of Science Education, 35*(5), 753-772. doi:10.1080/09500693.2011.577843
Goetz, T., Hall, N., Frenzel, A., & Pekrun, R. (2006). A hierarchical conceptualization of enjoyment in students. *Learning and Instruction, 16*(4), 323-338. doi:10.1016/j.learninstruc.2006.07.004
Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., . . . Preckel, F. (2014). ''My Questionnaire is Too Long!'' The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology, 39*(3), 188-205. doi:10.1016/j.cedpsych.2014.04.002
Gonida, E., & Leondari, A. (2011). Patterns of motivation among adolescents with biased and accurate self-efficacy beliefs. *International Journal of Educational Research, 50*(4), 209-220. doi:10.1016/j.ijer.2011.08.002
Gramzow, R., & Willard, G. (2006). Exaggerating Current and Past Performance: Motivated Self-Enhancement Versus Reconstructive Memory. *Personality and Social Psychology Bulletin, 32*(8), 1114-1125. doi:10.1177/0146167206288600
Gramzow, R., Elliot, A., Asher, E., & McGregor, H. (2003). Self-evaluation bias and academic performance: Some ways and some reasons why. *Journal of Research in Personality, 37*(2), 41-61. doi:10.1016/S0092-6566(02)00535-4
Gresham, F., Lane, K., MacMillan, D., Bocian, K., & Ward, S. (2000). Effects of Positive and Negative Illusory Biases: Comparisons Across Social and Academic Self-Concept Domains. *Journal of School Psychology, 38*(2), 151-175. doi:10.1016/S0022-4405(99)00042-4
Guay, F., Ratelle, C., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences, 20*(6), 644-653. doi:10.1016/j.lindif.2010.08.001
Gunderson, E., Ramirez, G., Levine, S., & Beilock, S. (2012). The Role of Parents and Teachers in the Development of Gender-Related Math Attitudes. *Sex Roles, 66*(3), 153-166. doi:10.1007/s11199-011-9996-2
Guo, J., Marsh, H., Morin, A., Parker, P., & Kaur, G. (2015). Directionality of the Associations of High School Expectancy-Value, Aspirations, and Attainment: A Longitudinal Study. *American Educational Research Journal, 52*(2), 371-402. doi:10.3102/0002831214565786
Guo, J., Nagengast, B., Marsh, H., Kelava, A., Gaspard, H., Brandt, H., . . . Trautwein, U. (2016). Probing the Unique Contributions of Self-Concept, Task Values, and Their Interactions Using Multiple Value Facets and Multiple Academic Outcomes. *AERA Open, 2*(1), 1-20. doi:10.1177/2332858415626884
Gutierrez, A., Schraw, G., Kuch, F., & Richmond, A. (2016). A two-process model of metacognitive monitoring: Evidence for general accuracy and error factors. *Learning and Instruction, 44*, 1-10. doi:10.1016/j.learninstruc.2016.02.006
Hacker, D., Bol, L., Horgan, D., & Rakow, E. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*(1), 160-170. doi:10.1037/0022-0663.92.1.160

Hadwin, A., Nesbit, J., Jamieson-Noel, D., Code, J., & Winne, P. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning, 2*(2-3), 107-124. doi:10.1007/s11409-007-9016-7

Haladyna, T., & Rodriguez, M. (2013). *Developing and Validating Test Items*. New York: Routledge.

Hampden-Thompson, G., & Bennett, J. (2013). Science Teaching and Learning Activities and Students' Engagement in Science. *International Journal of Science Education, 35*(8), 1325-1343. doi:10.1080/09500693.2011.608093

Hansford, B., & Hattie, J. (1982). The Relationship Between Self and Achievement/Performance Measures. *Review of Educational Research, 52*(1), 123-142. doi:10.3102/00346543052001123

Hansson, P., Rönnlund, M., Juslin, P., & Nilsson, L.-G. (2008). Adult age differences in the realism of confidence judgments: Overconfidence, format dependence, and cognitive predictors. *Psychology and Aging, 23*(3), 531-544. doi:10.1037/a0012782

Hanushek, E., & Kimko, D. (2000). Schooling, labor-force quality, and the growth of nations. *The American Economic Review, 90*(5), 1184-1208. doi:10.1257/aer.90.5.1184

Hanushek, E., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth, 17*(4), 267-321. doi:10.1007/s10887-012-9081-x

Harackiewicz, J., Rozek, C., Hulleman, C., & Hyde, J. (2012). Helping Parents to Motivate Adolescents in Mathematics and Science: An Experimental Test of a Utility-Value Intervention. *Psychological Science, 23*(8), 898-906. doi:10.1177/0956797611435530

Hardy, G. (2014). Academic Self-Concept: Modeling and Measuring for Science. *Research in Science Education, 44*(4), 549-579. doi:10.1007/s11165-013-9393-7

Harlen, W. (2005). Trusting teachers' judgement: research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education, 20*(3), 245-270. doi:10.1080/02671520500193744

Harter, S., Rumbaugh Whitesell, N., & Kowalski, P. (1992). Individual Differences in the Effects of Educational Transitions on Young Adolescent's Perceptions of Competence and Motivational Orientation. *American Educational Research Journal, 29*(4), 777-807. doi:10.3102/00028312029004777

He, Q., & Stockford, I. (2015). *Inter-Subject Comparability of Exam Standards in GCSE and A Level*. Coventry: Office of Qualifications and Examinations (Ofqual).

Heine, S. (2005). Where Is the Evidence for Pancultural Self-Enhancement? A Reply to Sedikides, Gaertner, and Toguchi (2003). *Journal of Personality and Social Psychology, 89*(4), 531-538. doi:10.1037/0022-3514.89.4.531

Heine, S., & Hamamura, T. (2007). In Search of East Asian Self-Enhancement. *Personality and Social Psychology Review, 11*(1), 4-27. doi:10.1177/1088868306294587

Heine, S., Lehman, D., Markus, H., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review, 106*(4), 766-794. doi:10.1037/0033-295X.106.4.766

Henson, J., Reise, S., & Kim, K. (2007). Detecting Mixtures From Structural Model Differences Using Latent Variable Mixture Modeling: A Comparison of Relative Model Fit Statistics. *Structural Equation Modeling, 14*(2), 202-226. doi:10.1080/10705510709336744

Hepper, E., Gramzow, R., & Sedikides, C. (2010). Individual Differences in Self-Enhancement and Self-Protection Strategies: An Integrative Analysis. *Journal of Personality, 78*(2), 781-814. doi:10.1111/j.1467-6494.2010.00633.x

Hepper, E., Sedikides, C., & Cai, H. (2013). Self-Enhancement and Self-Protection Strategies in China: Cultural Expressions of a Fundamental Human Motive. *Journal of Cross-Cultural Psychology, 44*(1), 5-23. doi:10.1177/0022022111428515

Hernandez-Martinez, P., Black, L., Williams, J., Davis, P., Pampaka, M., & Wake, G. (2008). Mathematics students' aspirations for higher education: class, ethnicity, gender and interpretative repertoire styles. *Research Papers in Education, 23*(2), 153-165. doi:10.1080/02671520802048687

Hodson, D. (1993). Re-thinking Old Ways: Towards A More Critical Approach To Practical Work In School Science. *Studies in Science Education, 22*(1), 85-142. doi:10.1080/03057269308560022

Holmegaard, H. T., Madsen, L. M., & Ulriksen, L. (2014). To Choose or Not to Choose Science: Constructions of desirable identities among young people considering a STEM higher education programme. *International Journal of Science Education, 36*(2), 186-215. doi:10.1080/09500693.2012.749362
Holmegaard, H. T., Ulriksen, L., & Madsen, L. M. (2014). The Process of Choosing What to Study: A Longitudinal Study of Upper Secondary Students' Identity Work When Choosing Higher Education. *Scandinavian Journal of Educational Research, 58*(1), 21-40. doi:10.1080/00313831.2012.696212
Holmegaard, H. T., Ulriksen, L., & Madsen, L. M. (2015). A Narrative Approach to Understand Students' Identities and Choices. In E. K. Henriksen, J. Dillon, & J. Ryder (Eds.), *Understanding Student Participation and Choice in Science and Technology Education* (pp. 31-42). Dordrecht: Springer. doi:10.1007/978-94-007-7793-4_3
Homer, M., Ryder, J., & Banner, I. (2014). Measuring determinants of post-compulsory participation in science: a comparative study using national data. *British Educational Research Journal, 40*(4), 610-636. doi:10.1002/berj.3106
Hong, H.-Y., & Lin-Siegler, X. (2012). How Learning About Scientists' Struggles Influences Students' Interest and Learning in Physics. *Journal of Educational Psychology, 104*(2), 469-484. doi:10.1037/a0026224
Hox, J. (2002). *Multilevel Analysis Techniques and Applications*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
Huang, C. (2011a). Achievement Goals and Achievement Emotions: A Meta-analysis. *Educational Psychology Review, 23*(3), 359-388. doi:10.1007/s10648-011-9155-x
Huang, C. (2011b). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology, 49*(5), 505-528. doi:10.1016/j.jsp.2011.07.001
Huff, J., & Nietfeld, J. (2009). Using strategy instruction and confidence judgments to improve metacognitive monitoring. *Metacognition and Learning, 4*(2), 161-176. doi:10.1007/s11409-009-9042-8
Huguet, P., Dumas, F., Marsh, H., Régner, I., Wheeler, L., Suls, J., . . . Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish–little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology, 97*(1), 156-170. doi:10.1037/a0015558
Huguet, P., Dumas, F., Monteil, J., & Genestoux, N. (2001). Social comparison choices in the classroom: further evidence for students' upward comparison tendency and its beneficial impact on performance. *European Journal of Social Psychology, 31*(5), 557-578. doi:10.1002/ejsp.81
Hulleman, C., & Harackiewicz, J. (2009). Promoting Interest and Performance in High School Science Classes. *Science, 326*(5958), 1410-1412. doi:10.1126/science.1177067
Hyde, J., Fennema, E., Ryan, M., Frost, L., & Hopp, C. (1990). Gender Comparisons of Mathematics Attitudes and Affect: A Meta-Analysis. *Psychology of Women Quarterly, 14*(3), 299-324. doi:10.1111/j.1471-6402.1990.tb00022.x
IBM. (2014). *IBM SPSS Missing Values 22*. IBM. Retrieved April 1, 2016, from ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/22.0/en/client/Manuals/IBM_SPSS_Missing_Values.pdf
Institute of Physics. (2013). *Closing Doors: Exploring gender and subject choice in schools*. London: Institute of Physics.
Institute of Physics. (2014). *Raising Aspirations in Physics: A review of research into barriers to STEM participation for students from disadvantaged backgrounds*. London: Institute of Physics.
Institution of Mechanical Engineers. (2014). *Five Tribes: Personalising Engineering Education*. London: Institution of Mechanical Engineers.
Ireson, J., & Hallam, S. (2009). Academic self-concepts in adolescence: Relations with achievement and ability grouping in schools. *Learning and Instruction, 19*(3), 201-213. doi:10.1016/j.learninstruc.2008.04.001
Jacobs, J., Lanza, S., Osgood, D. W., Eccles, J., & Wigfield, A. (2002). Changes in Children's Self-Competence and Values: Gender and Domain Differences across Grades One through Twelve. *Child Development, 73*(2), 509-527. doi:10.1111/1467-8624.00421
Jacobse, A., & Harskamp, E. (2012). Towards efficient measurement of metacognition in mathematical problem solving. *Metacognition Learning, 7*(2), 133-149. doi:10.1007/s11409-012-9088-x

Jansen, M., Lüdtke, O., & Ulrich, S. (2016). Evidence for a positive relation between interest and achievement: Examining between-person and within-person variation in five domains. *Contemporary Educational Psychology, 46*, 116-127. doi:10.1016/j.cedpsych.2016.05.004

Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology, 41*, 13-24. doi:10.1016/j.cedpsych.2014.11.002

Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: Multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences, 30*, 11-21. doi:10.1016/j.lindif.2013.12.003

Jansen, M., Schroeders, U., Lüdtke, O., & Marsh, H. (2015). Contrast and Assimilation Effects of Dimensional Comparisons in Five Subjects: An Extension of the I/E Model. *Journal of Educational Psychology, 107*(4), 1086-1101. doi:10.1037/edu0000021

Jenkins, E., & Nelson, N. (2005). Important but not for me: students' attitudes towards secondary school science in England. *Research in Science and Technological Education*, *23*(1), 41-57. doi:10.1080/02635140500068435

Jensen, F., & Henriksen, E. K. (2015). Short Stories of Educational Choice: In the Words of Science and Technology Students. In E. K. Henriksen, J. Dillon, & J. Ryder (Eds.), *Understanding Student Participation and Choice in Science and Technology Education* (pp. 135-151). Dordrecht: Springer. doi:10.1007/978-94-007-7793-4_9

Jiang, Y., Song, J., Lee, M., & Bong, M. (2014). Self-efficacy and achievement goals as motivational links between perceived contexts and achievement. *Educational Psychology, 34*(1), 92-117. doi:10.1080/01443410.2013.863831

Jurik, V., Gröschner, A., & Seidel, T. (2013). How student characteristics affect girls' and boys' verbal engagement in physics instruction. *Learning and Instruction, 23*, 33-42. doi:10.1016/j.learninstruc.2012.09.002

Jurik, V., Gröschner, A., & Seidel, T. (2014). Predicting students' cognitive learning activity and intrinsic learning motivation: How powerful are teacher statements, student profiles, and gender? *Learning and Individual Differences, 32*, 132-139. doi:10.1016/j.lindif.2014.01.005

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard–easy effect. *Psychological Review, 107*(2), 384-396. doi:10.1037/0033-295X.107.2.384

Kelemen, W., Frost, P., & Weaver, C. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition, 28*(1), 92-107. doi:10.3758/BF03211579

Khattab, N. (2015). Students' aspirations, expectations and school achievement: what really matters? *British Educational Research Journal, 41*(5), 731-748. doi:10.1002/berj.3171

Kimball, D. (1913). Practical Work in Science Teaching. *Science, 38*(970), 144-149. doi:10.1126/science.38.970.144

Kitayama, S., Markus, H. R., Matsumoto, H., & Norasakkunkit, V. (1997). Individual and collective processes in the construction of the self: Self-enhancement in the United States and self-criticism in Japan. *Journal of Personality and Social Psychology, 72*(6), 1245-1267. doi:10.1037/0022-3514.72.6.1245

Kjærnsli, M., & Lie, S. (2011). Students' Preference for Science Careers: International comparisons based on PISA 2006. *International Journal of Science Education, 33*(1), 121-144. doi:10.1080/09500693.2010.518642

Klar, Y., & Giladi, E. (1999). Are Most People Happier than their Peers, or Are They Just Happy? *Personality and Social Psychology Bulletin, 25*(5), 586-595. doi:10.1177/0146167299025005004

Kolb, D. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. New Jersey: Prentice-Hall.

Kolen, M., & Brennan, R. (2004). *Test Equating, Scaling, and Linking: Methods and Practices* (2nd ed.). New York: Springer.

Köller, O. (2001). Mathematical world views and achievement in advanced mathematics in Germany: findings from TIMSS population 3. *Studies in Educational Evaluation, 27*(1), 65-78. doi:10.1016/S0191-491X(01)00014-1

Köller, O., Baumert, J., & Schnabel, K. (2001). Does Interest Matter: The Relationship between Academic Interest and Achievement in Mathematics. *Journal for Research in Mathematics Education, 32*(5), 448-470. doi:10.2307/749801

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349-370. doi:10.1037/0096-3445.126.4.349

Koriat, A. (2012). The Self-Consistency Model of Subjective Confidence. *Psychological Review, 119*(1), 80-113. doi:10.1037/a0025648

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing Objective and Subjective Learning Curves: Judgments of Learning Exhibit Increased Underconfidence With Practice. *Journal of Experimental Psychology: General, 131*(2), 147-162. doi:10.1037//0096-3445.131.2.147

Krajc, M., & Ortmann, A. (2008). Are the unskilled really that unaware? An alternative explanation. *Journal of Economic Psychology, 29*(5), 724-738. doi:10.1016/j.joep.2007.12.006

Kramarski, B., Mevarech, Z., & Arami, M. (2002). The effects of metacognitive instruction on solving mathematical authentic tasks. *Educational Studies in Mathematics, 49*(2), 225-250. doi:10.1023/A:1016282811724

Krueger, J., & Mueller, R. (2002). Unskilled, Unaware, or Both? The Better-Than-Average Heuristic and Statistical Regression Predict Errors in Estimates of Own Performance. *Journal of Personality and Social Psychology, 82*(2), 180-188. doi:10.1037//0022-3514.82.2.180

Krueger, J., & Wright, J. (2010). Measurement of Self-Enhancement (and Self-Protection). In M. Alicke, & C. Sedikides (Eds.), *Handbook of Self-Enhancement and Self-Protection* (pp. 472-494). New York: Guilford Press.

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*(2), 221-232. doi:10.1037/0022-3514.77.2.221

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134. doi:10.1037/0022-3514.77.6.1121

Kuncel, N., Credé, M., & Thomas, L. (2005). The Validity of Self-Reported Grade Point Averages, Class Ranks, and Test Scores: A Meta-Analysis and Review of the Literature. *Review of Educational Research, 75*(1), 63-82. doi:10.3102/00346543075001063

Labuhn, A. S., Zimmerman, B., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: the influence of feedback and self-evaluative standards. *Metacognition and Learning, 5*(2), 173-194. doi:10.1007/s11409-010-9056-2

Lazarides, R., & Watt, H. (2015). Girls' and boys' perceived mathematics teacher beliefs, classroom learning environments and mathematical career intentions. *Contemporary Educational Psychology, 41*, 51-61. doi:10.1016/j.cedpsych.2014.11.005

Lee, J., Bong, M., & Kim, S.-I. (2014). Interaction between task values and self-efficacy on maladaptive achievement strategy use. *Educational Psychology, 34*(5), 538-560. doi:10.1080/01443410.2014.895296

Lee, W., Lee, M.-J., & Bong, M. (2014). Testing interest and self-efficacy as predictors of academic self-regulation and achievement. *Contemporary Educational Psychology, 39*(2), 86-99. doi:10.1016/j.cedpsych.2014.02.002

Leitch, S. (2006). *Prosperity for all in the global economy - world class skills*. London: Her Majesty's Stationery Office.

Lent, R., Lopez, F., & Bieschke, K. (1991). Mathematics Self-Efficacy: Sources and Relation to Science-Based Career Choice. *Journal of Counseling Psychology, 38*(4), 424-430. doi:10.1037/0022-0167.38.4.424

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20*(2), 159-183. doi:10.1016/0030-5073(77)90001-0

Lichtenstein, S., Fischhoff, B., & Phillips, L. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). Cambridge: Cambridge University Press.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22*(140), 1-55.
Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen-Ntoumani, C. (2012). Method Effects: The Problem With Negatively Versus Positively Keyed Items. *Journal of Personality Assessment, 94*(2), 196-204. doi:10.1080/00223891.2011.645936
Lippmann Kung, R., & Linder, C. (2007). Metacognitive activity in the physics student laboratory: is increased metacognition necessarily better? *Metacognition Learning, 2*(1), 41-56. doi:10.1007/s11409-007-9006-9
Little, R. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association, 83*(404), 1198-1202. doi:10.1080/01621459.1988.10478722
Lopez, F., Lent, R., Brown, S., & Gore, P. (1997). Role of social–cognitive expectations in high school students' mathematics-related interest and performance. *Journal of Counseling Psychology, 44*(1), 44-52. doi:10.1037/0022-0167.44.1.44
Lubbers, M., Kuyper, H., & van der Werf, M. (2009). Social comparison with friends versus non-friends. *European Journal of Social Psychology, 39*(1), 52-68. doi:10.1002/ejsp.475
Ma, X., & Kishor, N. (1997). Attitude Toward Self, Social Factors, and Achievement in Mathematics: A Meta-Analytic Review. *Educational Psychology Review, 9*(2), 89-120. doi:10.1023/A:1024785812050
Maas, C., & Hox, J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology, 1*(3), 86-92. doi:10.1027/1614-2241.1.3.86
Mabe, P., & West, S. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*(3), 280-296. doi:10.1037/0021-9010.67.3.280
Machts, N., Kaiser, J., Schmidt, F., & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review, 19*, 85-103. doi:10.1016/j.edurev.2016.06.003
MacKinnon, D., Kisbu-Sakarya, Y., & Gottschall, A. (2013). Developments in Mediation Analysis. In T. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology* (Vol. 2, pp. 338-360). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199934898.013.0016
Maki, R., Shields, M., Wheeler, A., & Zacchilli, T. L. (2005). Individual Differences in Absolute and Relative Metacomprehension Accuracy. *Journal of Educational Psychology, 97*(4), 723-731. doi:10.1037/0022-0663.97.4.723
Maltese, A., & Tai, R. (2010). Eyeballs in the Fridge: Sources of early interest in science. *International Journal of Science Education, 32*(5), 669-685. doi:10.1080/09500690902792385
Maltese, A., & Tai, R. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education, 95*(5), 877-907. doi:10.1002/sce.20441
Maltese, A., Melki, C., & Wiebke, H. (2014). The Nature of Experiences Responsible for the Generation and Maintenance of Interest in STEM. *Science Education, 98*(6), 937-962. doi:10.1002/sce.21132
Marsh, H. (1986a). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology, 22*(1), 37-49. doi:10.1037/0012-1649.22.1.37
Marsh, H. (1986b). Verbal and Math Self-Concepts: An Internal/External Frame of Reference Model. *American Educational Research Journal, 23*(1), 129-149. doi:10.3102/00028312023001129
Marsh, H. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79*(3), 280-295. doi:10.1037/0022-0663.79.3.280
Marsh, H. (1989). Age and sex effects in multiple dimensions of self-concept: preadolescence to early childhood. *Journal of Educational Psychology, 81*(3), 417-430. doi:10.1037/0022-0663.81.3.417
Marsh, H. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology, 82*(4), 623-636. doi:10.1037/0022-0663.82.4.623
Marsh, H., & Hau, K.-T. (2003). Big-Fish–Little-Pond Effect on Academic Self-Concept: A Cross-Cultural (26-Country) Test of the Negative Effects of Academically Selective Schools. *American Psychologist, 58*(5), 364-376. doi:10.1037/0003-066X.58.5.364

Marsh, H., & Köller, O. (2004). Unification of theoretical models of academic self-concept/achievement relations: Reunification of east and west German school systems after the fall of the Berlin Wall. *Contemporary Educational Psychology, 29*(3), 264-282. doi:10.1016/S0361-476X(03)00034-1

Marsh, H., & Parker, J. (1984). Determinants of Student Self-Concept: Is It Better to Be a Relatively Large Fish in a Small Pond Even if You Don't Learn to Swim as Well? *Journal of Personality and Social Psychology, 47*(1), 213-231. doi:10.1037/0022-3514.47.1.213

Marsh, H., & Shavelson, R. (1985). Self-Concept: Its Multifaceted, Hierarchical Structure. *Educational Psychologist, 20*(3), 107-123. doi:10.1207/s15326985ep2003_1

Marsh, H., & Yeung, A. S. (1998). Longitudinal Structural Equation Models of Academic Self-Concept and Achievement: Gender Differences in the Development of Math and English Constructs. *American Educational Research Journal, 35*(4), 705-738. doi:10.3102/00028312035004705

Marsh, H., Abduljabbar, A. S., Morin, A., Parker, P., Abdelfattah, F., Nagengast, B., & Abu-Hilal, M. (2015). The big-fish-little-pond effect: Generalizability of social comparison processes over two age cohorts from Western, Asian, and Middle Eastern Islamic countries. *Journal of Educational Psychology, 107*(1), 258-271. doi:10.1037/a0037485

Marsh, H., Byrne, B., & Shavelson, R. (1988). A Multifaceted Academic Self-Concept: Its Hierarchical Structure and Its Relation to Academic Achievement. *Journal of Educational Psychology, 80*(3), 366-380. doi:10.1037/0022-0663.80.3.366

Marsh, H., Craven, R., & Debus, R. (1999). Separation of Competency and Affect Components of Multiple Dimensions of Academic Self-Concept: A Developmental Perspective. *Merrill-Palmer Quarterly, 45*(4), 567-601.

Marsh, H., Kuyper, H., Seaton, M., Parker, P., Morin, A., Möller, J., & Abduljabbar, A. (2014). Dimensional comparison theory: an extension of the internal/external frame of reference effect on academic self-concept formation. *Contemporary Educational Psychology, 39*(4), 326-341. doi:10.1016/j.cedpsych.2014.08.003

Marsh, H., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., & Jansen, M. (2015). Dimensional Comparison Theory: Paradoxical relations between self-beliefs and achievements in multiple domains. *Learning and Instruction, 35*, 16-32. doi:10.1016/j.learninstruc.2014.08.005

Marsh, H., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic Self-Concept, Interest, Grades, and Standardized Test Scores: Reciprocal Effects Models of Causal Ordering. *Child Development, 76*(2), 397-416. doi:10.1111/j.1467-8624.2005.00853.x

Martin, A., & Debus, R. (1998). Self-reports of mathematics self-concept and educational outcomes: the roles of ego-dimensions and self-consciousness. *British Journal of Educational Psychology, 68*(4), 517-535. doi:10.1111/j.2044-8279.1998.tb01309.x

Martin, M., & Mullis, I. (Eds.). (2013). *Methods and Procedures in TIMSS and PIRLS 2011*. Chestnut Hill, Massachusetts: TIMSS & PIRLS International Study Centre. Retrieved from http://timssandpirls.bc.edu/methods/

Masyn, K. (2013). Latent Class Analysis and Finite Mixture Modeling. In T. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology* (Vol. II, pp. 551-611). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199934898.013.0025

McCrone, T., Morris, M., & Walker, M. (2005). *Student Choices at Key Stage 3 – Literature Review. DfES Research Report RW68*. London: Department for Education and Skills.

Mellors-Bourne, R., Connor, H., & Jackson, C. (2011). *STEM Graduates in Non-STEM Jobs*. London: Department for Business, Innovation and Skills.

Merkle, E., Sieck, W., & van Zandt, T. (2008). Response error and processing biases in confidence judgment. *Journal of Behavioral Decision Making, 21*(4), 428-448. doi:10.1002/bdm.597

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. doi:10.1037/0003-066X.50.9.741

Mezulis, A., Abramson, L., Hyde, J., & Hankin, B. (2004). Is There a Universal Positivity Bias in Attributions? A Meta-Analytic Review of Individual, Developmental, and Cultural Differences in the Self-Serving Attributional Bias. *Psychological Bulletin, 130*(5), 711-747. doi:10.1037/0033-2909.130.5.711

Millar, R. (1998). Rhetoric and reality: What practical work in science is really for. In J. Wellington (Ed.), *Practical work in school science. Which way now?* (pp. 16-31). London: Routledge.
Möller, J., & Köller, O. (2001). Dimensional comparisons: An experimental approach to the internal/external frame of reference model. *Journal of Educational Psychology, 93*(4), 826-835. doi:10.1037/0022-0663.93.4.826
Möller, J., & Marsh, H. (2013). Dimensional Comparison Theory. *Psychological Review, 120*(3), 544-560. doi:10.1037/a0032459
Möller, J., Pohlmann, B., Köller, O., & Marsh, H. (2009). A Meta-Analytic Path Analysis of the Internal/External Frame of Reference Model of Academic Achievement and Academic Self-Concept. *Review of Educational Research, 79*(3), 1129-1167. doi:10.3102/0034654309337522
Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. (2011). The Reciprocal Internal/External Frame of Reference Model: An Integration of Models of Relations Between Academic Achievement and Self-Concept. *American Educational Research Journal, 48*(6), 1315-1346. doi:10.3102/0002831211419649
Morgan, S., Gelbgiser, D., & Weeden, K. (2013). Feeding the pipeline: Gender, occupational plans, and college major selection. *Social Science Research, 42*(4), 989-1005. doi:10.1016/j.ssresearch.2013.03.008
Muis, K., Winne, P., & Jamieson-Noel, D. (2007). Using a multitrait-multimethod analysis to examine conceptual similarities of three self-regulated learning inventories. *British Journal of Educational Psychology, 77*(1), 177-195. doi:10.1348/000709905X90876
Mujtaba, T., & Reiss, M. J. (2013). What Sort of Girl Wants to Study Physics After the Age of 16? Findings from a Large-scale UK Survey. *International Journal of Science Education, 35*(17), 2979-2998. doi:10.1080/09500693.2012.681076
Mujtaba, T., & Reiss, M. J. (2014). A survey of psychological, motivational, family and perceptions of physics education factors that explain 15-year-old students' aspirations to study physics in post-compulsory English schools. *International Journal of Science and Mathematics Education, 12*(2), 371-393. doi:10.1007/s10763-013-9404-1
Mullis, I., Martin, M., Ruddock, G., O'Sullivan, C., & Preuschoff, C. (2009). *TIMSS 2011 Assessment Frameworks*. Chestnut Hill, Massachusetts: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
Mullola, S., Hintsanen, M., Jokela, M., Lipsanen, J., Alatupa, S., Ravaja, N., & Keltikangas-Järvinen, L. (2014). Associations Between Teacher-Rated Versus Self-Rated Student Temperament and School Achievement. *Scandinavian Journal of Educational Research, 58*(2), 147-172. doi:10.1080/00313831.2012.725094
Multon, K., Brown, S., & Lent, R. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30-38. doi:10.1037/0022-0167.38.1.30
Murphy, P. K., & Alexander, P. (2000). A Motivated Exploration of Motivation Terminology. *Contemporary Educational Psychology, 25*(1), 3-53. doi:10.1006/ceps.1999.1019
Murphy, P., & Whitelegg, E. (2006). *Institute of Physics Report: Girls in the Physics Classroom: A Review of the Research on the Participation of Girls in Physics*. London: Institute of Physics.
Musu-Gillette, L., Wigfield, A., Harring, J., & Eccles, J. (2015). Trajectories of change in students' self-concepts of ability and values in math and college major choice. *Educational Research and Evaluation, 21*(4), 343-370. doi:10.1080/13803611.2015.1057161
Muthén, L., & Muthén, B. (2012). *Mplus User's Guide* (7th ed.). Los Angeles, California: Muthén & Muthén.
Nagengast, B., & Marsh, H. (2011). The negative effect of school-average ability on science self-concept in the UK, the UK countries and the world: the Big-Fish-Little-Pond-Effect for PISA 2006. *Educational Psychology, 31*(5), 629-659. doi:10.1080/01443410.2011.586416
Nagengast, B., Marsh, H., Scalas, L. F., Xu, M., Hau, K.-T., & Trautwein, U. (2011). Who Took the "×" out of Expectancy-Value Theory? A Psychological Mystery, a Substantive-Methodological Synergy, and a Cross-National Generalization. *Psychological Science, 22*(8), 1058-1066. doi:10.1177/0956797611415540
Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and Course Selection in Upper Secondary Education: Effects of academic self-concept and

intrinsic value. *Educational Research and Evaluation, 12*(4), 323-345. doi:10.1080/13803610600765687
Nagy, G., Watt, H., Eccles, J., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The Development of Students' Mathematics Self-Concept in Relation to Gender: Different Countries, Different Trajectories? *Journal of Research on Adolescence, 20*(2), 482-506. doi:10.1111/j.1532-7795.2010.00644.x
Narciss, S., Koerndle, H., & Dresel, M. (2011). Self-evaluation accuracy and satisfaction with performance: Are there affective costs or benefits of positive self-evaluation bias? *International Journal of Educational Research, 50*(4), 230-240. doi:10.1016/j.ijer.2011.08.004
Niepel, C., Brunner, M., & Preckel, F. (2014a). Achievement goals, academic self-concept, and school grades in mathematics: Longitudinal reciprocal relations in above average ability secondary school students. *Contemporary Educational Psychology, 39*(4), 301-313. doi:10.1016/j.cedpsych.2014.07.002
Niepel, C., Brunner, M., & Preckel, F. (2014b). The Longitudinal Interplay of Students' Academic Self-Concepts and Achievements Within and Across Domains: Replicating and Extending the Reciprocal Internal/External Frame of Reference Model. *Journal of Educational Psychology, 106*(4), 1170-1191. doi:10.1037/a0036307
Nylund, K., Asparouhov, T., & Muthén, B. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling, 14*(4), 535-569. doi:10.1080/10705510701575396
OECD [Organisation for Economic Co-operation and Development]. (2006a). *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*. Paris: OECD Publishing. doi:10.1787/9789264026407-en
OECD [Organisation for Economic Co-operation and Development]. (2006b). *PISA released items - science*. Paris: OECD.
OECD [Organisation for Economic Co-operation and Development]. (2007). *PISA 2006 Science Competencies for Tomorrow's World. Volume 1: Analysis*. Paris: OECD Publishing. doi:10.1787/9789264040014-en
OECD [Organisation for Economic Co-operation and Development]. (2009a). *PISA 2006 Technical Report*. Paris: OECD Publishing. doi:10.1787/9789264048096-en
OECD [Organisation for Economic Co-operation and Development]. (2009b). *PISA Data Analysis Manual: SPSS* (2nd ed.). OECD Publishing.
OECD [Organisation for Economic Co-operation and Development]. (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD Publishing. doi:10.1787/9789264190511-en
OECD [Organisation for Economic Co-operation and Development]. (2014). *PISA 2012 Technical Report*. Paris: OECD Publishing.
OECD [Organisation for Economic Co-operation and Development]. (2015a). *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*. Paris: OECD Publishing. doi:10.1787/9789264229945-en
OECD [Organisation for Economic Co-operation and Development]. (2015b). *Universal Basic Skills: What Countries Stand to Gain*. Paris: OECD Publishing. doi:10.1787/9789264234833-en
Osborne, J., & Dillon, J. (2008). *Science Education in Europe: Critical Reflections*. London: The Nuffield Foundation.
Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049-1079. doi:10.1080/0950069032000032199
Pajares, F., & Graham, L. (1999). Self-Efficacy, Motivation Constructs, and Mathematics Performance of Entering Middle School Students. *Contemporary Educational Psychology, 24*, 124-139. doi:10.1006/ceps.1998.0991
Pajares, F., & Miller, M. D. (1995). Mathematics Self-Efficacy and Mathematics Performances: The Need for Specificity of Assessment. *Journal of Counseling Psychology, 42*(2), 190-198. doi:10.1037/0022-0167.42.2.190
Parker, P., Marsh, H., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology, 34*(1), 29-48. doi:10.1080/01443410.2013.797339

Parker, P., Marsh, H., Morin, A., Seaton, M., & Van Zanden, B. (2015). If one goes up the other must come down: Examining ipsative relationships between math and English self-concept trajectories across high school. *British Journal of Educational Psychology, 85*(2), 172-191. doi:10.1111/bjep.12050

Parker, P., Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U., & Eccles, J. (2012). Achievement, Agency, Gender, and Socioeconomic Background as Predictors of Postschool Choices: A Multicontext Study. *Developmental Psychology, 48*(6), 1629-1642. doi:10.1037/a0029167

Paulhus, D. (1991). Measurement and Control of Response Bias. In J. Robinson, P. Shaver, & L. Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes* (pp. 17-59). London: Academic Press.

Paulhus, D., & Dubois, P. (2014). Application of the Overclaiming Technique to Scholastic Assessment. *Educational and Psychological Measurement, 74*(6), 975-990. doi:10.1177/0013164414536184

Paulhus, D., Harms, P., Bruce, M. N., & Lysy, D. (2003). The Over-Claiming Technique: Measuring Self-Enhancement Independent of Ability. *Journal of Personality and Social Psychology, 84*(4), 890-904. doi:10.1037/0022-3514.84.4.890

Paulick, I., Watermann, R., & Nückles, M. (2013). Achievement goals and school achievement: The transition to different school tracks in secondary school. *Contemporary Educational Psychology, 38*(1), 75-86. doi:10.1016/j.cedpsych.2012.10.003

Peugh, J., & Enders, C. (2004). Missing Data in Educational Research: A Review of Reporting Practices and Suggestions for Improvement. *Review of Educational Research, 74*(4), 525-556. doi:10.3102/00346543074004525

Phillips, J., & Gully, S. (1997). Role of goal orientation, ability, need for achievement, and locus of control in the self-efficacy and goal-setting process. *Journal of Applied Psychology, 82*(5), 792-802. doi:10.1037/0021-9010.82.5.792

Pieschl, S. (2009). Metacognitive calibration—an extended conceptualization and potential applications. *Metacognition and Learning, 4*(1), 3-31. doi:10.1007/s11409-008-9030-4

Pintrich, P. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research, 31*(6), 459-470. doi:10.1016/S0883-0355(99)00015-4

Pintrich, P. (2004). A Conceptual Framework for Assessing Motivation and Self-Regulated Learning in College Students. *Educational Psychology Review, 16*(4), 385-407. doi:10.1007/s10648-004-0006-x

Pintrich, P., Smith, D., Garcia, T., & Mckeachie, W. (1993). Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3), 801-813. doi:10.1177/0013164493053003024

Pintrich, P., Smith, D., Garcia, T., & Wilbert, M. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire*. Ann Arbor, Michigan: National Center for Research to Improve Postsecondary Teaching and Learning, The University of Michigan. Retrieved October 22, 2013, from http://eric.ed.gov/?id=ED338122

Plante, I., de la Sablonnière, R., Aronson, J., & Théorêt, M. (2013). Gender stereotype endorsement and achievement-related outcomes: The role of competence beliefs and task values. *Contemporary Educational Psychology, 38*(3), 225-235. doi:10.1016/j.cedpsych.2013.03.004

Potvin, P., & Hasni, A. (2014). Analysis of the Decline in Interest Towards School Science and Technology from Grades 5 Through 11. *Journal of Science Education and Technology, 23*(6), 784-802. doi:10.1007/s10956-014-9512-x

Praetorius, A.-K., Berner, V.-D., Zeinz, H., Scheunpflug, A., & Dresel, M. (2013). Judgment Confidence and Judgment Accuracy of Teachers in Judging Self-Concepts of Students. *The Journal of Educational Research, 106*(1), 64-76. doi:10.1080/00220671.2012.667010

Preckel, F., Niepel, C., Schneider, M., & Brunner, M. (2013). Self-concept in adolescence: A longitudinal study on reciprocal effects of self-perceptions in academic and social domains. *Journal of Adolescence, 36*(6), 1165-1175. doi:10.1016/j.adolescence.2013.09.001

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). *GLLAMM Manual*. Berkeley, California: University of California, Berkeley. Retrieved June 24, 2014, from http://biostats.bepress.com/ucbbiostat/paper160

Regan, E., & DeWitt, J. (2015). Attitudes, Interest and Factors Influencing STEM Enrolment Behaviour: An Overview of Relevant Literature. In E. K. Henriksen, J. Dillon, & J. Ryder (Eds.), *Understanding Student Participation and Choice in Science and Technology Education* (pp. 63-88). Dordrecht: Springer. doi:10.1007/978-94-007-7793-4_5

Reiss, M. J., Hoyles, C., Mujtaba, T., Riazi-Farzad, B., Rodd, M., Simon, S., & Stylianidou, F. (2011). Understanding Participation Rates in Post-16 Mathematics and Physics: Conceptualising and Operationalising the UPMAP Project. *International Journal of Science and Mathematics Education, 9*(2), 273-302. doi:10.1007/s10763-011-9286-z

Rhodes, J., Roffman, J., Reddy, R., & Fredriksen, K. (2004). Changes in self-esteem during the middle school years: a latent growth curve study of individual and contextual influences. *Journal of School Psychology, 42*(3), 243-261. doi:10.1016/j.jsp.2004.04.001

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological Correlates of University Students' Academic Performance: A Systematic Review and Meta-Analysis. *Psychological Bulletin, 138*(2), 353-387. doi:10.1037/a0026838

Rinne, L., & Mazzocco, M. (2014). Knowing Right From Wrong In Mental Arithmetic Judgments: Calibration Of Confidence Predicts The Development Of Accuracy. *PLoS ONE, 9*(7), e98663. doi:10.1371/journal.pone.0098663

Roberts, G. (2002). *SET for success: The supply of people with science, technology, engineering and mathematics skills. The report of Sir Gareth Roberts' Review*. London: The Stationery Office.

Robins, R., & Beer, J. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology, 80*(2), 340-352. doi:10.1037/0022-3514.80.2.340

Rogosa, D., & Willett, J. (1983). Demonstrating the Reliability of the Difference Score in the Measurement of Change. *Journal of Educational Measurement, 20*(4), 335-343. doi:10.1111/j.1745-3984.1983.tb00211.x

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92*(3), 726-748. doi:10.1037/0033-2909.92.3.726

Rose, D., & O'Reilly, K. (1998). *The ESRC Review of Government Social Classifications*. London: Office for National Statistics.

Rose, D., Pevalin, D., & O'Reilly, K. (2005). *The National Statistics Socio-economic Classification: Origins, Development and Use*. Houndmills: Palgrave Macmillan.

Rosenzweig, E., & Wigfield, A. (2016). STEM Motivation Interventions for Adolescents: A Promising Start, but Further to Go. *Educational Psychologist, 51*(2), 146-163. doi:10.1080/00461520.2016.1154792

Rost, D., Sparfeldt, J., Dickhauser, O., & Schilling, S. (2005). Dimensional comparisons in subject-specific academic self-concepts and achievements: A quasi-experimental approach. *Learning and Instruction, 15*(6), 557-570. doi:10.1016/j.learninstruc.2005.08.003

Rotter, J. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*(1), 1-28. doi:10.1037/h0092976

Royal Society. (2006). *A degree of concern? UK first degrees in science, technology and mathematics*. London: The Royal Society.

Royal Society. (2008a). *A higher degree of concern*. London: The Royal Society.

Royal Society. (2008b). *Science and mathematics education, 14–19: A 'state of the nation' report on the participation and attainment of 14–19 year olds in science and mathematics in the UK, 1996–2007*. London: The Royal Society.

Royal Society. (2014). *Vision for science and mathematics education*. London: The Royal Society.

Rozek, C., Hyde, J., Svoboda, R., Hulleman, C., & Harackiewicz, J. (2015). Gender Differences in the Effects of a Utility-Value Intervention to Help Parents Motivate Adolescents in Mathematics and Science. *Journal of Educational Psychology, 107*(1), 195-206. doi:10.1037/a0036981

Rubin, D. (1976). Inference and missing data. *Biometrika, 63*(3), 581-592. doi:10.1093/biomet/63.3.581

Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons. doi:10.1002/9780470316696

Rubin, D. (1996). Multiple Imputation after 18+ Years. *Journal of the American Statistical Association, 91*(434), 473-489. doi:10.1080/01621459.1996.10476908

Ryan, R., & Deci, E. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54-67. doi:10.1006/ceps.1999.1020

Rytkönen, K., Aunola, K., & Nurmi, J.-E. (2007). Do Parents' Causal Attributions Predict the Accuracy and Bias in their Children's Self-Concept of Maths Ability? A longitudinal study. *Educational Psychology, 27*(6), 771-788. doi:10.1080/01443410701309316

Sabate, R., Harris, A., & Staff, J. (2011). Ambition Gone Awry: The Long-Term Socioeconomic Consequences of Misaligned and Uncertain Ambitions in Adolescence. *Social Science Quarterly, 92*(4), 959-977. doi:10.1111/j.1540-6237.2011.00799.x

Sáinz, M., & Upadyaya, K. (2016). Accuracy and bias in Spanish secondary school students' self-concept of math ability: The influence of gender and parental educational level. *International Journal of Educational Research, 77*, 26-36. doi:10.1016/j.ijer.2016.02.009

Samuelson, P. (1942). A Note on Alternative Regressions. *Econometrica, 10*(1), 80-83. doi:10.2307/1907024

Schaefer, P., Williams, C., Goodie, A., & Campbell, W. K. (2004). Overconfidence and the Big Five. *Journal of Research in Personality, 38*(5), 473-480. doi:10.1016/j.jrp.2003.09.010

Schellings, G. (2011). Applying learning strategy questionnaires: problems and possibilities. *Metacognition and Learning, 6*(2), 91-109. doi:10.1007/s11409-011-9069-5

Schellings, G., van Hout-Wolters, B., Veenman, M., & Meijer, J. (2013). Assessing metacognitive activities: the in-depth comparison of a task-specific questionnaire with think-aloud protocols. *European Journal of Psychology of Education, 28*(3), 963-990. doi:10.1007/s10212-012-0149-y

Schoon, I. (2001). Teenage job aspirations and career attainment in adulthood: A 17-year follow-up study of teenagers who aspired to become scientists, health professionals, or engineers. *International Journal of Behavioral Development, 25*(2), 124-132. doi:10.1080/01650250042000186

Schraw, G., & Dennison, R. (1994). Assessing Metacognitive Awareness. *Contemporary Educational Psychology, 19*(4), 460-475. doi:10.1006/ceps.1994.1033

Schraw, G., Dunkle, M., Bendixen, L., & Roedel, T. D. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology, 87*(3), 433-444. doi:10.1037/0022-0663.87.3.433

Schraw, G., Kuch, F., & Gutierrez, A. (2013). Measure for measure: Calibrating ten commonly used calibration scores. *Learning and Instruction, 24*, 48-57. doi:10.1016/j.learninstruc.2012.08.007

Schraw, G., Kuch, F., Gutierrez, A., & Richmond, A. (2014). Exploring a Three-Level Model of Calibration Accuracy. *Journal of Educational Psychology, 106*(4), 1192-1202. doi:10.1037/a0036653

Schroeder, C., Scott, T., Tolson, H., Huang, T.-Y., & Lee, Y.-H. (2007). A Meta-Analysis of National Research: Effects of Teaching Strategies on Student Achievement in Science in the United States. *Journal of Research in Science Teaching, 44*(10), 1436-1460. doi:10.1002/tea.20212

Schwab, S., & Hessels, M. (2015). Achievement Goals, School Achievement, Self-Estimations of School Achievement, and Calibration in Students With and Without Special Education Needs in Inclusive Education. *Scandinavian Journal of Educational Research, 59*(4), 461-477. doi:10.1080/00313831.2014.932304

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annuals of Statistics, 6*(2), 461-464. doi:10.1214/aos/1176344136

Seaton, M., Marsh, H., & Craven, R. (2009). Earning Its Place as a Pan-Human Theory: Universality of the Big-Fish-Little-Pond Effect Across 41 Culturally and Economically Diverse Countries. *Journal of Educational Psychology, 101*(2), 403-419. doi:10.1037/a0013838

Seaton, M., Marsh, H., Dumas, F., Huguet, P., Monteil, J.-M., Régner, I., . . . Wheeler, L. (2008). In search of the big fish: Investigating the coexistence of the big-fish-little-pond effect with the positive effects of upward comparisons. *British Journal of Social Psychology, 47*(1), 73-103. doi:10.1348/014466607X202309

Seaton, M., Parker, P., Marsh, H., Craven, R., & Yeung, A. S. (2014). The reciprocal relations between self-concept, motivation and achievement: juxtaposing academic self-concept and achievement goal orientations for mathematics success. *Educational Psychology, 34*(1), 49-72. doi:10.1080/01443410.2013.825232

Sedikides, C., & Alicke, M. (2012). Self-Enhancement and Self-Protection Motives. In R. Ryan (Ed.), *The Oxford Handbook of Human Motivation* (pp. 303-322). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780195399820.013.0017

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural Self-Enhancement. *Journal of Personality and Social Psychology, 84*(1), 60-79. doi:10.1037/0022-3514.84.1.60

Seidel, T. (2006). The role of student characteristics in studying micro teaching–learning environments. *Learning Environments Research, 9*(3), 253-271. doi:10.1007/s10984-006-9012-x

Sfard, A., & Prusak, A. (2005). Telling Identities: In Search of an Analytic Tool for Investigating Learning as a Culturally Shaped Activity. *Educational Researcher, 34*(4), 14-22. doi:10.3102/0013189X034004014

Shavelson, R., Hubner, J., & Stanton, G. (1976). Self-Concept: Validation of Construct Interpretations. *Review of Educational Research, 46*(3), 407-441. doi:10.2307/1170010

Sheldrake, R. (2016a). Confidence as motivational expressions of interest, utility, and other influences: Exploring under-confidence and over-confidence in science students at secondary school. *International Journal of Educational Research, 76*, 50-65. doi:10.1016/j.ijer.2015.12.001

Sheldrake, R. (2016b). Differential predictors of under-confidence and over-confidence for mathematics and science students in England. *Learning and Individual Differences, 49*, 305-313. doi:10.1016/j.lindif.2016.05.009

Sheldrake, R. (2016c). Students' intentions towards studying science at upper-secondary school: the differential effects of under-confidence and over-confidence. *International Journal of Science Education, 38*(8), 1256-1277. doi:10.1080/09500693.2016.1186854

Sheldrake, R., Mujtaba, T., & Reiss, M. J. (2014). Calibration of self-evaluations of mathematical ability for students in England aged 13 and 15, and their intentions to study non-compulsory mathematics after age 16. *International Journal of Educational Research, 64*, 49-61. doi:10.1016/j.ijer.2013.10.008

Sheldrake, R., Mujtaba, T., & Reiss, M. J. (2015). Students' intentions to study non-compulsory mathematics: the importance of how good you think you are. *British Educational Research Journal, 41*(3), 462-488. doi:10.1002/berj.3150

Silver, A., & Rushton, B. (2008). Primary-school children's attitudes towards science, engineering and technology and their images of scientists and engineers. *Education 3–13, 36*(1), 51-67. doi:10.1080/03004270701576786

Simms, L. (2008). Classical and Modern Methods of Psychological Scale Construction. *Social and Personality Psychology Compass, 2*(1), 414-433. doi:10.1111/j.1751-9004.2007.00044.x

Simpkins, S., & Davis-Kean, P. (2005). The intersection between self-concepts and values: Links between beliefs and choices in high school. *New Directions for Child and Adolescent Development, 2005*(110), 31-47. doi:10.1002/cd.148

Simpkins, S., Davis-Kean, P., & Eccles, J. (2006). Math and Science Motivation: A Longitudinal Examination of the Links Between Choices and Beliefs. *Developmental Psychology, 42*(1), 70-83. doi:10.1037/0012-1649.42.1.70

Simpson, R., & Oliver, J. S. (1990). A summary of major influences on attitude toward and achievement in science among adolescent students. *Science Education, 74*(1), 1-18. doi:10.1002/sce.3730740102

Sjaastad, J. (2012). Sources of Inspiration: The role of significant persons in young people's choice of science in higher education. *International Journal of Science Education, 34*(10), 1615-1636. doi:10.1080/09500693.2011.590543

Skaalvik, E., & Hagtvet, K. (1990). Academic achievement and self-concept: An analysis of causal predominance in a developmental perspective. *Journal of Personality and Social Psychology, 58*(2), 292-307. doi:10.1037/0022-3514.58.2.292

Skaalvik, E., & Rankin, R. (1990). Math, Verbal, and General Academic Self-Concept: The Internal/External Frame of Reference Model and Gender Differences in Self-Concept Structure. *Journal of Educational Psychology, 82*(3), 546-554. doi:10.1037/0022-0663.82.3.546

Skaalvik, E., & Rankin, R. (1995). A Test of the Internal/External Frame of Reference Model at Different Levels of Math and Verbal Self-Perception. *American Educational Research Journal, 32*(1), 161-184. doi:10.3102/00028312032001161

Skaalvik, E., Federici, R., & Klassen, R. (2015). Mathematics achievement and self-efficacy: Relations with motivation for mathematics. *International Journal of Educational Research, 72*, 129-136. doi:10.1016/j.ijer.2015.06.008

Skaalvik, S., & Skaalvik, E. (2004). Gender Differences in Math and Verbal Self-Concept, Performance Expectations, and Motivation. *Sex Roles, 50*(3/4), 241-252. doi:10.1023/B:SERS.0000015555.40976.e6

Sliter, K., & Zickar, M. (2014). An IRT Examination of the Psychometric Functioning of Negatively Worded Personality Items. *Educational and Psychological Measurement, 74*(2), 214-226. doi:10.1177/0013164413504584

Smith, E. (2011). Women into science and engineering? Gendered participation in higher education STEM subjects. *British Educational Research Journal, 37*(6), 993-1014. doi:10.1080/01411926.2010.515019

Snijders, T., & Bosker, R. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd ed.). London: SAGE Publications.

Soll, J., & Klayman, J. (2004). Overconfidence in Interval Estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 299-314. doi:10.1037/0278-7393.30.2.299

Sperling, R., Howard, B., Staley, R., & DuBoi, N. (2004). Metacognition and Self-Regulated Learning Constructs. *Educational Research and Evaluation, 10*(2), 117-139. doi:10.1076/edre.10.2.117.27905

Sperling, R., Richmond, A., Ramsay, C., & Klapp, M. (2012). The Measurement and Predictive Ability of Metacognition in Middle School Learners. *The Journal of Educational Research, 105*(1), 1-7. doi:10.1080/00220671.2010.514690

Stankov, L., Lee, J., & Paek, I. (2009). Realism of Confidence Judgments. *European Journal of Psychological Assessment, 25*(2), 123-130. doi:10.1027/1015-5759.25.2.123

StataCorp. (2013a). *STATA Multiple-Imputation Reference Manual release 13*. College Station, Texas: Stata Press. Retrieved July 1, 2015, from http://www.stata.com/manuals13/mi.pdf

StataCorp. (2013b). *STATA Power and Sample-Size Reference Manual release 13*. College Station, Texas: Stata Press. Retrieved July 1, 2015, from http://www.stata.com/manuals13/pss.pdf

Stevens, T., Wang, K., Olivárez Jr., A., & Hamman, D. (2007). Use of Self-perspectives and their Sources to Predict the Mathematics Enrollment Intentions of Girls and Boys. *Sex Roles, 56*(5-6), 351-363. doi:10.1007/s11199-006-9180-2

Stone, N. (2000). Exploring the Relationship between Calibration and Self-Regulated Learning. *Educational Psychology Review, 12*(4), 437-475. doi:10.1023/A:1009084430926

Strand, S. (2007). *Minority Ethnic Pupils in the Longitudinal Study of Young People in England (LSYPE). Research Report DCSF-RR002*. Nottingham: DfES Publications.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743-762. doi:10.1037/a0027627

Taconis, R., & Kessels, U. (2009). How Choosing Science depends on Students' Individual Fit to 'Science Culture'. *International Journal of Science Education, 31*(8), 1115-1132. doi:10.1080/09500690802050876

Tai, R., Qi Liu, C., Maltese, A., & Fan, X. (2006). Planning early for careers in science. *Science, 312*(5777), 1143-1145. doi:10.1126/science.1128690

Tan, E., Calabrese Barton, A., Kang, H., & O'Neill, T. (2013). Desiring a career in STEM-related fields: How middle school girls articulate and negotiate identities-in-practice in science. *Journal of Research in Science Teaching, 50*(10), 1143-1179. doi:10.1002/tea.21123

Taskinen, P., Schütte, K., & Prenzel, M. (2013). Adolescents' motivation to select an academic science-related career: the role of school factors, individual interest, and science self-concept. *Educational Research and Evaluation, 19*(8), 717-733. doi:10.1080/13803611.2013.853620

Thijs, J., Verkuyten, M., & Helmond, P. (2010). A Further Examination of the Big-Fish–Little-Pond Effect: Perceived Position in Class, Class Size, and Gender

Comparisons. *Sociology of Education, 83*(4), 333-345. doi:10.1177/0038040710383521
Torgerson, C., Gorard, S., Low, G., Ainsworth, H., See, B. H., & Wright, K. (2008). *What are the factors that promote high post-16 participation of many minority ethnic groups? A focused review of the UK-based aspirations literature*. London: Evidence for Policy and Practice Information and Co-ordinating Centre.
Trautwein, U., Marsh, H., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the Multiplicative Term in Modern Expectancy–Value Theory: A Latent Interaction Modeling Study. *Journal of Educational Psychology, 104*(3), 763-777. doi:10.1037/a0027470
Tripney, J., Newman, M., Bangpan, M., Niza, C., MacKintosh, M., & Sinclair, J. (2010). *Factors Influencing Young People (Aged 14-19) in Education about STEM Subject Choices: A systematic review of the UK literature*. London: Wellcome Trust.
Tröbst, S., Kleickmann, T., Lange-Schubert, K., Rothkopf, A., & Möller, K. (2016). Instruction and Students' Declining Interest in Science: An Analysis of German Fourth- and Sixth-Grade Classrooms. *American Educational Research Journal, 53*(1), 162-193. doi:10.3102/0002831215618662
Turner, S., & Ireson, G. (2010). Fifteen pupils' positive approach to primary school science: when does it decline? *Educational Studies, 36*(2), 119-141. doi:10.1080/03055690903148662
Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen, & K. Elliot (Eds.), *But What Does It Mean? The Use of Effect Sizes in Educational Research* (pp. 55-66). Slough: National Foundation for Educational Research.
Urhahne, D., Chao, S.-H., Florineth, M. L., Luttenberger, S., & Paechter, M. (2011). Academic self-concept, learning motivation, and test anxiety of the underestimated student. *British Journal of Educational Psychology, 81*(1), 161-177. doi:10.1348/000709910X504500
Usher, E., & Pajares, F. (2008a). Self-Efficacy for Self-Regulated Learning: A Validation Study. *Educational and Psychological Measurement, 68*(2), 443-463. doi:10.1177/0013164407308475
Usher, E., & Pajares, F. (2008b). Sources of Self-Efficacy in School: Critical Review of the Literature and Future Directions. *Review of Educational Research, 78*(4), 751-796. doi:10.3102/0034654308321456
Usher, E., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology, 34*(1), 89-101. doi:10.1016/j.cedpsych.2008.09.002
van Dinther, M., Dochy, F., & Segers, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review, 6*(2), 95-108. doi:10.1016/j.edurev.2010.10.003
Vermunt, J., & Magidson, J. (2013). *Technical Guide for Latent GOLD 5.0: Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations. Retrieved June 1, 2015, from http://www.statisticalinnovations.com/technicalsupport/LGtechnical.pdf
Vidal Rodeiro, C. (2007). *A level subject choice in England: patterns of uptake and factors affecting subject preferences*. Cambridge: Cambridge Assessment.
Viljaranta, J., Tolvanen, A., Aunola, K., & Nurmi, J.-E. (2014). The Developmental Dynamics between Interest, Self-concept of Ability, and Academic Performance. *Scandinavian Journal of Educational Research, 58*(6), 734-756. doi:10.1080/00313831.2014.904419
Walker, I., & Zhu, Y. (2013). *Impact of University Degrees on the Lifecycle of Earnings: Some Further Analysis*. London: Department for Business, Innovation and Skills.
Wang, M.-T. (2012). Educational and Career Interests in Math: A Longitudinal Examination of the Links Between Classroom Environment, Motivational Beliefs, and Interests. *Developmental Psychology, 48*(6), 1643-1657. doi:10.1037/a0027247
Wang, M.-T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy-value perspective to understand individual and gender differences in STEM fields. *Developmental Review, 33*(4), 304-340. doi:10.1016/j.dr.2013.08.001
Wang, X. (2013). Why Students Choose STEM Majors: Motivation, High School Learning, and Postsecondary Context of Support. *American Educational Research Journal, 50*(5), 1081-1121. doi:10.3102/0002831213488622

Watt, H. (2006). The Role of Motivation in Gendered Educational and Occupational Trajectories Related to Maths. *Educational Research and Evaluation, 12*(4), 305-322. doi:10.1080/13803610600765562
Weems, G., Onwuegbuzie, A., & Collins, K. (2006). The Role of Reading Comprehension in Responses to Positively and Negatively Worded Items on Rating Scales. *Evaluation and Research in Education, 19*(1), 3-20. doi:10.1080/09500790608668322
Weems, G., Onwuegbuzie, A., Schreiber, J., & Eggers, S. (2003). Characteristics of respondents who respond differently to positively and negatively worded items on rating scales. *Assessment and Evaluation in Higher Education, 28*(6), 587-606. doi:10.1080/0260293032000130234
Wehrens, M., Kuyper, H., Dijkstra, P., Buunk, A., & van der Werf, M. (2010). The long-term effect of social comparison on academic performance. *European Journal of Social Psychology, 40*(7), 1158-1171. doi:10.1002/ejsp.706
West, R., & Stanovich, K. (1997). The domain specificity and generality of overconfidence: Individual differences in performance estimation bias. *Psychonomic Bulletin & Review, 4*(3), 387-392. doi:10.3758/BF03210798
Wetzel, E., Carstensen, C., & Böhnke, J. (2013). Consistency of extreme response style and non-extreme response style across traits. *Journal of Research in Personality, 47*(2), 178-189. doi:10.1016/j.jrp.2012.10.010
Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review, 30*(1), 1-35. doi:10.1016/j.dr.2009.12.001
Wigfield, A., & Eccles, J. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12*(3), 265-310. doi:10.1016/0273-2297(92)90011-P
Wigfield, A., & Eccles, J. (2000). Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology, 25*(1), 68-81. doi:10.1006/ceps.1999.1015
Wigfield, A., Eccles, J., Mac Iver, D., Reuman, D., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology, 27*(4), 552-565. doi:10.1037/0012-1649.27.4.552
Willard, G., & Gramzow, R. (2008). Exaggeration in memory: Systematic distortion of self-evaluative information under reduced accessibility. *Journal of Experimental Social Psychology, 44*(2), 246-259. doi:10.1016/j.jesp.2007.04.012
Willard, G., & Gramzow, R. (2009). Beyond Oversights, Lies, and Pies in the Sky: Exaggeration as Goal Projection. *Personality and Social Psychology Review, 35*(4), 477-492. doi:10.1177/0146167208329631
Willett, J. (1988). Questions and Answers in the Measurement of Change. *Review of Research in Education, 15*(1), 345-422. doi:10.3102/0091732X015001345
Williams, J., & Choudry, S. (2016). Mathematics capital in the educational field: Bourdieu and beyond. *Research in Mathematics Education, 18*(1), 3-21. doi:10.1080/14794802.2016.1141113
Wilson, R., Beaven, R., May-Gillings, M., Hay, G., & Stevens, J. (2014). *Working Futures 2012-2022*. London: UK Commission for Employment and Skills.
Winne, P. (1995). Inherent details in self-regulated learning. *Educational Psychologist, 30*(4), 173-187. doi:10.1207/s15326985ep3004_2
WISE [Women in Science, Technology and Engineering]. (2014). *"Not for people like me?" Under-represented groups in science, technology and engineering*. Bradford: WISE.
Wong, B. (2012). Identifying with Science: A case study of two 13-year-old 'high achieving working class' British Asian girls. *International Journal of Science Education, 34*(1), 43-65. doi:10.1080/09500693.2010.551671
Wong, B. (2015). Careers "From" but not "in" science: Why are aspirations to be a scientist challenging for minority ethnic students? *Journal of Research in Science Teaching, 52*(7), 979-1002. doi:10.1002/tea.21231
Wong, B. (2016). Minority Ethnic Students and Science Participation: a Qualitative Mapping of Achievement, Aspiration, Interest and Capital. *Research in Science Education, 46*(1), 113-127. doi:10.1007/s11165-015-9466-x
Woolley, E. (1941). The Method of Minimized Areas as a Basis for Correlation Analysis. *Econometrica, 9*(1), 38-62. doi:10.2307/1907173

Yaniv, I., Yates, J. F., & Smith, J. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin, 110*(3), 611-617. doi:10.1037/0033-2909.110.3.611

Yates, J. F. (1990). *Judgment and decision making*. Englewood Cliffs, New Jersey: Prentice-Hall.

Young, J., & Mroczek, D. (2003). Predicting intraindividual self-concept trajectories during adolescence. *Journal of Adolescence, 26*(5), 586-600. doi:10.1016/S0140-1971(03)00058-7

Yu, H., Jiang, S., & Land, K. (2015). Multicollinearity in hierarchical linear models. *Social Science Research, 53*, 118-136. doi:10.1016/j.ssresearch.2015.04.008

Yule, G. U. (1912). On the Methods of Measuring Association Between Two Attributes. *Journal of the Royal Statistical Society, 75*(6), 579-652. doi:10.2307/2340126

Zeldin, A., & Pajares, F. (2000). Against the Odds: Self-Efficacy Beliefs of Women in Mathematical, Scientific, and Technological Careers. *American Educational Research Journal, 37*(1), 215-246. doi:10.3102/00028312037001215

Zeldin, A., Britner, S., & Pajares, F. (2008). A Comparative Study of the Self-Efficacy Beliefs of Successful Men and Women in Mathematics, Science, and Technology Careers. *Journal of Research in Science Teaching, 45*(9), 1036-1058. doi:10.1002/tea.20195

Zell, E., & Krizan, Z. (2014). Do People Have Insight Into Their Abilities? A Metasynthesis. *Perspectives on Psychological Science, 9*(2), 111-125. doi:10.1177/1745691613518075

Zhai, J., Jocz, J. A., & Tan, A.-L. (2014). 'Am I Like a Scientist?': Primary children's images of doing science in school. *International Journal of Science Education, 36*(4), 553-576. doi:10.1080/09500693.2013.791958

Zheng, X., & Rabe-Hesketh, S. (2007). Estimating parameters of dichotomous and ordinal item response models with gllamm. *The Stata Journal, 7*(3), 313-333.

Zimmerman, B. (2000). Attaining self-regulation: A social–cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). Burlington, MA: Elsevier Academic Press.

Zimmerman, B., & Martinez-Pons, M. (1986). Development of a Structured Interview for Assessing Student Use of Self-Regulated Learning Strategies. *American Educational Research Journal, 23*(4), 614-628. doi:10.3102/00028312023004614

Zimmerman, B., & Martinez-Pons, M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology, 80*(3), 284-290. doi:10.1037/0022-0663.80.3.284

Zimmerman, B., & Moylan, A. (2009). Self-regulation: Where metacognition and motivation intersect. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299-315). New York: Routledge.

Zimmerman, B., & Schunk, D. (2011). Self-Regulated Learning and Performance: An Introduction and an Overview. In B. Zimmerman, & D. Schunk (Eds.), *Handbook of Self-Regulation of Learning and Performance* (pp. 1-12). Abingdon: Routledge.

Appendix 1: Item/factor comparability across surveys

Table A1.1: Item-level survey comparability for measuring self-concept beliefs

	Example source/reference items		
2014/2015 survey items	PISA 2006	TIMSS 2011	PISA 2012
'I usually do well in science'	'I can usually give good answers to test questions on science topics'	'I usually do well in science'	-
'I have always been good at science'	-	-	-
'I get good grades in science'	-	-	'I get good grades in science'
'I understand even the most difficult science work'	'Learning advanced science topics would be easy for me'	'I am good at working out difficult science problems'	'In my science class, I understand even the most difficult work'
'I learn things quickly in science'	'I learn science topics quickly'	'I learn things quickly in science'	'I learn science quickly'
-	-	'Science is not one of my strengths'	'I am just not good at science'
-	'Science topics are easy for me'	-	-
-	'When I am being taught science, I can understand the concepts very well'	-	-
-	'I can easily understand new ideas in science'	-	-

Notes: PISA 2012 measured students' attitudes and beliefs related to mathematics; the listed items have been re-phrased to science for illustration.

Table A1.2: Item-level survey comparability for measuring interest value

	Example source/reference items		
2014/2015 survey items	PISA 2006	TIMSS 2011	PISA 2012
‘I look forward to my science lessons’	-	-	‘I look forward to my science lessons’
‘I am interested in the things I learn in science’	‘I am interested in learning about science’	‘I learn many interesting things in science’	‘I am interested in the things I learn in science’
‘I enjoy learning science’	‘I enjoy acquiring new knowledge in science’	‘I enjoy learning science’	‘I enjoy reading about science’
‘I like science’	‘I like reading about science’	‘I like science’	-
‘I like biology’	-	-	-
‘I like chemistry’	-	-	-
‘I like physics’	-	-	-
-	‘I generally have fun when I am learning science topics’	-	-
-	‘I am happy doing science problems’	-	‘I do science because I enjoy it’
-	-	‘I wish I did not have to study science’	-
-	-	‘Science is boring’	-

Notes: PISA 2012 measured students’ attitudes and beliefs related to mathematics; the listed items have been re-phrased to science for illustration.

Table A1.3: Item-level survey comparability for measuring utility value

2014/2015 survey items	Example source/reference items PISA 2006	TIMSS 2011	PISA 2012
'Making an effort in science is worth it because it will help me in the work that I want to do later on'	'Making an effort in science is worth it because this will help me in the work I want to do later on'	-	'Making an effort in science is worth it because it will help me in the work that I want to do later on'
'Learning science is worthwhile for me because it will improve my career prospects'	'Studying science is worthwhile for me because what I learn will improve my career prospects'	-	'Learning science is worthwhile for me because it will improve my career prospects'
'Science is an important subject for me because I need it for what I want to study later on'	'What I learn in science is important for me because I need this for what I want to study later on'	-	'Science is an important subject for me because I need it for what I want to study later on'
'I will learn many things in science that will help me get a job'	'I will learn many things in science that will help me get a job'	-	'I will learn many things in science that will help me get a job'
'I need to do well in science to get the job I want'	-	'I need to do well in science to get the job I want'	-
'I need to do well in science to get into the university of my choice'	-	'I need to do well in science to get into the university of my choice'	-
'I think learning science will help me in my daily life'	-	'I think learning science will help me in my daily life'	-
'I need science to learn other school subjects'	-	'I need science to learn other school subjects'	-
'I would like a job that involves using science'	-	'I would like a job that involves using science'	-
'Science is important to me'	-	-	-
'It is important to do well in science'	-	'It is important to do well in science'	-
-	'I study science because I know it is useful for me'	-	-

*Notes: PISA 2012 measured students' attitudes and beliefs related to mathematics; the listed items have been re-phrased to science for illustration. As highlighted in the methods (**Section 5.3**), some items (e.g. 'It is important to do well in science') appeared potentially similar to those covering personal value, but empirically loaded onto the utility value* *factor. Additionally, utility value aimed to measure the indirect or extrinsic benefits, importance, or value associated with science or studying science: items sourced from TIMSS (e.g. 'I would like a job that involves using science'; Martin & Mullis, 2013) aimed to measure preferences towards using scientific but transferable skills in a job not necessarily within science; the item potentially appeared similar to science intentions, but still loaded onto the utility value factor. Preliminary analysis highlighted that similar predictive coefficients were given regardless of whether the item was included or omitted from the factor.*

Table A1.4: Item-level survey comparability for measuring personal value

	Example source/reference items		
2014/2015 survey items	PISA 2006	TIMSS 2011	PISA 2012
'Science is important to me personally'	'Science is very relevant to me'	-	-
'Thinking scientifically is an important part of who I am'	-	-	-
'Being able to do science helps me show other people who I am'	-	-	-
-	'Some concepts in science help me see how I relate to other people'	-	-
-	'I will use science in many ways when I am an adult'	-	-
-	'I find that science helps me to understand the things around me'	-	-
-	'When I leave school there will be many opportunities for me to use science'	-	-

Notes: Personal value was referenced less from PISA/TIMSS and more from discrete studies (e.g. Conley, 2012; Trautwein, et al., 2012; Wigfield & Eccles, 2000) and broadly following theoretical aspects/dimensions of personal value or 'science identity' (e.g. the personal importance of science, science as an inherent aspect of personal identity, and science as a means to convey personal identity to other people; Carlone & Johnson, 2007).

Appendix 1.1: Item-level measurement in PISA 2006

The PISA 2006 items were mainly measured through agreement scales ('strongly disagree', 'disagree', 'agree', and 'strongly agree') (OECD, 2009a). As highlighted in **Section 5.2**, the OECD's intended/theorised factor composition, as below, was empirically confirmed for the students in England in PISA 2006.

Science intentions: 'I would like to work in a career involving science'; 'I would like to study science after secondary school; 'I would like to spend my life doing advanced science; 'I would like to work on science projects as an adult'.

Background factors (used as separate indicators/items): Gender (1=boy); Books at home ('How many books are there in your home'); Parent(s) working in science (1=yes); Mothers' education; Fathers' education.

Task-score (PV1): covered in detail within the OECD documentation (OECD, 2009a).

Self-concept: 'Learning advanced science topics would be easy for me'; 'I can usually give good answers to test questions on science topics'; 'I learn science topics quickly; 'Science topics are easy for me'; 'When I am being taught science I can understand the concepts very well'; 'I can easily understand new ideas in science'.

Self-efficacy (areas) ('How easy do you think it would be for you to perform the following tasks on your own?'): 'Recognise the science question that underlies a newspaper report on a health issue'; 'Explain why earthquakes occur more frequently in some areas than in others'; 'Describe the role of antibiotics in the treatment of disease'; 'Identify the science question associated with the disposal of garbage'; 'Predict how changes to an environment will affect the survival of certain species'; 'Interpret the scientific information provided on the labelling of food items'; 'Discuss how new evidence can lead you to change your understanding about the possibility of life on Mars'; 'Identify the better of two explanations for the formation of acid rain'.

Interest (various areas) ('How much interest do you have in learning about the following science topics?'): 'Topics in physics'; 'Topics in chemistry'; 'The biology of plants'; 'Human biology'; 'Topics in astronomy'; 'Topics in geology'; 'Ways scientists design experiments'; 'What is required for scientific explanations'.

Interest value: 'I generally have fun when I am learning science topics'; 'I like reading about broad science'; 'I am happy doing science problems'; 'I enjoy acquiring new knowledge in science'; 'I am interested in learning about broad science'.

Utility value: 'Making an effort in my science subjects is worth it because this will help me in the work I want to do later on'; 'What I learn in my science subjects is important for me because I need this for what I want to study later on'; 'I study science because I know it is useful for me'; 'Studying my science subjects is worthwhile for me because what I learn will improve my career prospects'; 'I will learn many things in my science subjects that will help me get a job'.

Personal value: 'Some concepts in science help me see how I relate to other people'; 'I will use science in many ways when I am an adult'; 'Science is very relevant to me'; 'I find that science helps me to understand the things around me'; 'When I leave school there will be many opportunities for me to use science'.

General value: 'Advances in science and technology usually improve people's living conditions'; 'Science is important for helping us to understand the natural world'; 'Advances in science and technology usually help improve the economy'; 'Science is valuable to society'; 'Advances in science and technology usually bring social benefits'.

Science activities ('How often do you do these things?'): 'Watch TV programmes about science'; 'Borrow or buy books on science topics'; 'Visit web sites about science topics'; 'Listen to radio programmes about advances in science'; 'Read science magazines or science articles in newspapers'; 'Attend a science club'.

School career preparation: 'The subjects available at my school provide students with the basic skills and knowledge for a science-related career'; 'The science subjects at my school provide students with the basic skills and knowledge for many different careers'; 'The subjects I study provide me with the basic skills and knowledge for a science-related career'; 'My teachers equip me with the basic skills and knowledge I need for a science-related career'.

School career information ('How informed are you about these topics?'): 'Science-related careers that are available in the job market'; 'Where to find information about science-related careers'; 'The steps a student needs to take if they want a science-related career'; 'Employers or companies that hire people to work in science-related careers'.

Teaching, interaction ('When learning science topics at school, how often do the following activities occur?'): 'Students are given opportunities to explain their ideas'; 'The lessons involve students' opinions about the topics'; 'There is a class debate or discussion'; 'The students have discussions about the topics'.

Teaching, activities ('When learning science topics at school, how often do the following activities occur?'): 'Students spend time in the laboratory doing practical experiments'; 'Students are required to design how a science question could be investigated in the laboratory'; 'Students are asked to draw conclusions from an experiment they have conducted'; 'Students do experiments by following the instructions of the teacher'.

Teaching, investigations ('When learning science topics at school, how often do the following activities occur?'): 'Students are allowed to design their own experiments'; 'Students are given the chance to choose their own investigations'; 'Students are asked to do an investigation to test out their own ideas'.

Teaching, applications ('When learning science topics at school, how often do the following activities occur?'): 'The teacher explains how a science idea can be applied to a number of different phenomena (e.g. the movement of objects, substances with similar properties)'; 'The teacher uses science to help students understand the world outside school'; 'The teacher clearly explains the relevance of science concepts to our lives'; 'The teacher uses examples of technological application to show how science is relevant to society'.

Appendix 1.2: Item-level measurement in the 2014/2015 survey

Following the questionnaire development (**Section 5.3**), which balanced covering multiple items/dimensions sourced or adapted from various prior studies (e.g. PISA 2006, TIMSS 2011, etc.) with empirical indicators of internal consistency and reliability, given the particular sample of students, the various factors were defined as summarised in **Section 5.3** and detailed below. For example, the factors covering the expectancy-value 'subjective task values' (interest value, utility value, personal value, and cost value) were broadly measured through items from both PISA and TIMSS surveys, together with other sources, with duplicate items removed and phrasing adapted (when necessary) to be clear and contextually relevant (e.g. Conley, 2012; Martin & Mullis, 2013; OECD, 2013; Trautwein, et al., 2012; Wigfield & Eccles, 2000).

As highlighted in **Section 5.3**, preliminary analysis confirmed that the results were insensitive to small changes in factor composition (i.e. the same conclusions would be drawn regarding relative coefficient magnitudes), such as measuring personal value or cost value with two or three items.

The questionnaire items mainly used agreement scales ('strongly disagree', 'disagree', 'slightly disagree', 'slightly agree', 'agree', and 'strongly agree'); depending on the item phrasing, categories were reverse-scored when necessary so that high item/factor values consistently indicated a positive experience or belief (e.g. doing well, being interested, the absence of anxiety, etc.).

Science intentions: 'I intend to study science at A-Level'; 'I intend to study science at university'; 'I am planning on pursuing a career that involves a lot of science'.

Background factors (used as separate indicators/items): Year; Gender (1=boy); Ethnicity; Books at home; Parent(s) working in science (1=yes); Mothers' education; Fathers' education.

Task-score: see the reproduced questionnaire for the particular items (**Appendix 3**).

Task-confidence: see the reproduced questionnaire for the particular items (**Appendix 3**).

Current grade: 'What overall grade/level have you got so far this year in science?'.

Self-concept: 'I usually do well in science'; 'I have always been good at science'; 'I get good grades in science'; 'I understand even the most difficult science work.'; 'I learn things quickly in science'.

Self-efficacy: 'What grade do you think you will be able to get at GCSE (or equivalent) science?'; 'What grade do you think you would be able to get if you studied your best science subject at A-Level?'.

Interest value: 'I look forward to my science lessons'; 'I am interested in the things I learn in science'; 'I enjoy learning science'; 'I like science'; 'I like biology'; 'I like chemistry'; 'I like physics'.

Utility value: 'Making an effort in science is worth it because it will help me in the work that I want to do later on'; 'Learning science is worthwhile for me because it will improve my career prospects'; 'Science is an important subject for me because I need it for what I want to study later on'; 'I will learn many things in science that will help me get a job'; 'I need to do well in science to get the job I want'; 'I need to do well in science to get into the university of my choice'; 'I think learning science will help me in my daily life'; 'I need science to learn other school subjects'; 'I would like a job that involves using science'; 'Science is important to me'; 'It is important to do well in science'.

Personal value: 'Science is important to me personally'; 'Thinking scientifically is an important part of who I am'; 'Being able to do science helps me show other people who I am'.

Cost value (absence of): 'I have to give up a lot to do well in science'; 'Success in science means that I need to give up other activities I enjoy'; 'I have to sacrifice a lot of free time to be good at science'; 'I have to invest a lot of time to get good grades in science'.

Orientation, mastery: 'I aim to understand and learn the material in science'.

Orientation, performance: 'I aim to perform better than other students in science'.

Perceived control: 'People can generally improve their ability in science'; 'I can improve my ability in science'; 'If I put in enough effort I can succeed in science'; 'Whether or not I do well in science is completely up to me'; 'If I wanted to, I could do well in science'.

Perceived control (exams): 'I do badly in science whether or not I study for my exams'; 'No matter whether or not I do my best in science, it will not improve my grades';

‘It is not worth practicing science for a test because I will come off badly again anyway’; ‘I accomplish almost nothing in science of what I intend to do’.

Study strategy, self-regulation: ‘During science class time I often miss important points because I'm thinking of other things’; ‘When studying for science, I make up questions to help focus my studying’; ‘When I become confused about something I'm studying for science class, I go back and try to figure it out’; ‘If science course materials are difficult to understand, I change the way I approach the material’; ‘Before I study new science course material thoroughly, I often skim it to see how it is organized’; ‘I ask myself questions to make sure I understand the material I have been studying earlier in science class’; ‘I try to change the way I study in order to fit the science course requirements and the teacher's teaching style’; ‘I often find that I have been studying for science class but don't know what it was all about’; ‘I try to think through a science topic and decide what I am supposed to learn from it rather than just reading it over when studying for science’; ‘When studying for this science course I try to determine which concepts I don't understand well’; ‘When I study for my science class, I set goals for myself in order to direct my activities in each study period’; ‘If I get confused taking notes in science class, I make sure I sort it out afterwards’.

Study strategy, control: ‘When I study for a science test, I try to work out what the most important parts to learn are’; ‘When I study science, I try to figure out which concepts I still have not understood properly’; ‘When I study science, I start by working out exactly what I need to learn’; ‘When I cannot understand something in science, I always search for more information to clarify the problem’.

Study strategy, memorisation: ‘When I study for a science test, I learn as much as I can off by heart’; ‘When I study science, I go over some problems so often that I feel as if I could solve them in my sleep’; ‘In order to remember the method for solving a science problem, I go through examples again and again’; ‘When I study science, I make myself check to see if I remember the work I have already done’.

Study strategy, elaboration: ‘When I study for a science test, I try to understand new concepts by relating them to things I already know’; ‘When I study science, I think of new ways to get the answer’; ‘When I study science, I try to relate the work to things I have learnt in other subjects’; ‘I think about how the science I have learnt can be used in everyday life’.

Anxiety (absence of): ‘Science makes me confused and nervous’; ‘I often worry that it will be difficult for me in science classes’; ‘I get very tense when I have to do science work’; ‘I feel helpless when doing a science problem’; ‘I worry that I will get poor grades in science’.

Mastery norms (good grade): ‘What grade do you think people need to get in order to be ‘good’ at science?’.

Subject-comparisons: ‘Science is harder for me than any other subject’.

Peer-comparisons: ‘Science is harder for me than for many of my classmates’.

Social persuasions (praise): ‘I have been praised for my ability in science’; ‘My science teacher tells me I am good at science’; ‘My science teacher thinks I can do well in science work with difficult materials’.

Vicarious experiences: ‘When I see how another student solves a science problem, I can see myself solving the problem in the same way’.

Norms/influence (friends): ‘Most of my friends do well in science’; ‘Most of my friends work hard at science’; ‘My friends enjoy taking science tests’.

Norms/influence (parents): ‘My parents believe it's important for me to study science’; ‘My parents believe that science is important for my career’; ‘My parents like science’.

Teacher perceptions: ‘I know what my science teacher expects me to do’; ‘My science teacher is easy to understand’; ‘I am interested in what my science teacher says’; ‘My science teacher gives me interesting things to do’; ‘My science teacher gives me feedback on my strengths and weaknesses in science’; ‘My science teacher tells us what is expected of us when we get a test, quiz or assignment’; ‘My science teacher tells us what we have to learn’; ‘My science teacher tells me what I need to do to become better in science’.

Teacher/school careers/events: ‘My science teacher tells me about careers and jobs in science’; ‘Other people at my school tell me about careers and jobs in science’; ‘My school does special activities, talks, events, or visits related to science (inside or outside of lessons)’.

Appendix 2: England 2014/2015 survey: sampled schools

Table A2.1a: England 2014/2015 (Years 9, 10, 11): contextualisation (2012/2013 data), school admissions policies

	England		Sampling frame		Invited		Participated	
Admissions policy	N	%	N	%	N	%	N	%
Comprehensive	2902	55.4%	2902	70.4%	269	85.7%	9	75.0%
Modern	134	2.6%	134	3.2%	13	4.1%	0	.0%
Selective	164	3.1	164	4.0%	17	5.4%	2	16.7%
Not applicable / other	783	14.9%	42	1.0%	1	.3%	0	.0%
Missing	1255	24.0%	883	21.4%	14	4.5%	1	8.3%
Total	5238	100.0%	4125	100.0%	314	100.0%	12	100.0%

Notes: Due to continual changes in school status, not all of the schools in the original sampling frame (4125 schools, from 2012/2013 data) and the invited schools (314 schools, from 2012/2013 data) could be matched against subsequent data.

Table A2.1b: England 2014/2015 (Years 9, 10, 11): contextualisation (2014/2015 data), school admissions policies

	England		Sampling frame		Invited		Participated	
Admissions policy	N	%	N	%	N	%	N	%
Comprehensive	3008	57.8%	2787	71.8%	265	86.0%	9	75.0%
Modern	121	2.3%	121	3.1%	13	4.2%	0	.0%
Selective	163	3.1%	163	4.2%	17	5.5%	2	16.7%
Not applicable / other	771	14.8%	1	.0%	0	.0%	0	.0%
Missing	1141	21.9%	808	20.8%	13	4.2%	1	8.3%
Total	5204	100.0%	3880	100.0%	308	100.0%	12	100.0%

Notes: Due to continual changes in school status, not all of the schools in the original sampling frame (4125 schools, from 2012/2013 data) and the invited schools (314 schools, from 2012/2013 data) could be matched against subsequent data.

Table A2.2a: England 2014/2015 (Years 9, 10, 11): contextualisation (2012/2013 data), school gender admissions policies

	England		Sampling frame		Invited		Participated	
Admissions policy	N	%	N	%	N	%	N	%
Boys	326	6.2%	241	5.8%	22	7.0%	3	25.0%
Girls	431	8.2%	421	10.2%	28	8.9%	1	8.3%
Mixed	4475	85.4%	3457	83.8%	264	84.1%	8	66.7%
Missing	6	.1%	6	.1%	0	.0%	0	.0%
Total	326	6.2%	4125	100.0%	314	100.0%	12	100.0%

Notes: Due to continual changes in school status, not all of the schools in the original sampling frame (4125 schools, from 2012/2013 data) and the invited schools (314 schools, from 2012/2013 data) could be matched against subsequent data.

Table A2.2b: England 2014/2015 (Years 9, 10, 11): contextualisation (2014/2015 data), school gender admissions policies

	England		Sampling frame		Invited		Participated	
Admissions policy	N	%	N	%	N	%	N	%
Boys	305	5.9%	224	5.8%	22	7.1%	3	25.0%
Girls	406	7.8%	390	10.1%	27	8.8%	1	8.3%
Mixed	4492	86.3%	3266	84.2%	259	84.1%	8	66.7%
Missing	1	.0%	0	.0%	0	.0%	0	.0%
Total	5204	100.0%	3880	100.0%	308	100.0%	12	100.0%

Notes: Due to continual changes in school status, not all of the schools in the original sampling frame (4125 schools, from 2012/2013 data) and the invited schools (314 schools, from 2012/2013 data) could be matched against subsequent data.

Table A2.3a: England 2014/2015 (Years 9, 10, 11): contextualisation (2012/2013 data), school admissions policies by gender admissions policies

	England			Sampling frame			Invited			Participated		
Admissions policy	Boys	Girls	Mixed	Boys	Girls	Mixed	Boys	Girls	Mixed	Boys	Girls	Mixed
Comprehensive	97	144	2661	97	144	2661	14	17	238	1	1	7
	2.4%	3.6%	66.9%	3.0%	4.4%	82.2%	4.7%	5.7%	79.3%	9.1%	9.1%	63.6%
Modern	12	16	106	12	16	106	3	1	9	0	0	0
	.3%	.4%	2.7%	.4%	.5%	3.3%	1.0%	.3%	3.0%	.0%	.0%	.0%
Selective	59	62	43	59	62	43	5	9	3	2	0	0
	1.5%	1.6%	1.1%	1.8%	1.9%	1.3%	1.7%	3.0%	1.0%	18.2%	.0%	.0%
Not applicable / other	44	2	731	0	0	36	0	0	1	0	0	0
	1.1%	.1%	18.4%	.0%	.0%	1.1%	.0%	.0%	.3%	.0%	.0%	.0%

Notes: Percentages sum to 100% for each of the main categories (e.g. England, the sampling frame, etc.). Due to continual changes in school status, not all of the schools in the original sampling frame (4125 schools, from 2012/2013 data) and the invited schools (314 schools, from 2012/2013 data) could be matched against subsequent data.

Table A2.3b: England 2014/2015 (Years 9, 10, 11): contextualisation (2014/2015 data), school admissions policies by gender admissions policies

	England			Sampling frame			Invited			Participated		
Admissions policy	Boys	Girls	Mixed	Boys	Girls	Mixed	Boys	Girls	Mixed	Boys	Girls	Mixed
Comprehensive	91	139	2777	85	136	2566	14	16	235	1	1	7
	2.2%	3.4%	68.4%	2.8%	4.4%	83.5%	4.7%	5.4%	79.7%	9.1%	9.1%	63.6%
Modern	11	14	96	11	14	96	3	1	9	0	0	0
	.3%	.3%	2.4%	.4%	.5%	3.1%	1.0%	.3%	3.1%	.0%	.0%	.0%
Selective	59	61	43	59	61	43	5	9	3	2	0	0
	1.5%	1.5%	1.1%	1.9%	2.0%	1.4%	1.7%	3.1%	1.0%	18.2%	.0%	.0%
Not applicable / other	41	2	728	0	0	1	0	0	0	0	0	0
	1.0%	.0%	17.9%	.0%	.0%	.0%	.0%	.0%	.0%	.0%	.0%	.0%

Notes: Percentages sum to 100% for each of the main categories (e.g. England, the sampling frame, etc.). Due to continual changes in school status, not all of the schools in the original sampling frame (4125 schools, from 2012/2013 data) and the invited schools (314 schools, from 2012/2013 data) could be matched against subsequent data.

Appendix 3: England 2014/2015 survey: questionnaire

The questionnaire has been reproduced within the format of this thesis; the typeface and size, page size (margins) and pagination, and the overall layout (including the size of response fields) differs from the original.

England 2014/2015 (Years 9, 10, 11): 10-task questionnaire

What do you think about science?

About this research

I am Richard Sheldrake from the Institute of Education at the University of London. I am exploring what students across the country think about science, and I would like you to take part in my research!

This questionnaire is anonymous (it does not ask for your name or for any identifying information) and confidential (your individual answers will only be seen by me). When I report on the results, I will analyse everyone's responses together and will not identify anyone.

If you have any questions about the research or want to find out more (including the results), you can contact me by email at rsheldrake@ioe.ac.uk

About this questionnaire

This questionnaire asks what you think about science, what you like or dislike, what you think about science lessons and other areas, and also about your general approaches to learning. I also have a few science puzzles, and I'm keen to know what you think and whether you can do them.

You may be studying science subjects separately, such as in classes for biology, chemistry, and physics, or you might study the areas together. **For this questionnaire, please try to think about science generally.** There are also questions where you can tell me about the separate areas of science.

Throughout the questionnaire, please tick the items that best represent you or what you think. Please only tick one item per question. You can leave any question blank if you don't want to answer it.

Thank you for your help!

About you

What school are you in?

____

☐ I am a boy.	☐ I am a girl.

What year are you in?

☐ Year 7	☐ Year 8	☐ Year 9	☐ Year 10	☐ Year 11	☐ Year 12	☐ Year 13

How would you describe your background or ethnicity?

☐ White British	☐ Indian
☐ White Irish	☐ Pakistani
☐ White European	☐ Bangladeshi
☐ Any other white background	☐ Any other Asian background
☐ Black African	☐ White and Asian
☐ Black Caribbean	☐ White and Black African
☐ Any other black background	☐ White and Black Caribbean
☐ Chinese	☐ Any other mixed background
☐ Any other East Asian background	☐ Any other ethnic group

How many books are there in your home? (Please include electronic books, but do not count magazines, newspapers, or your schoolbooks.)

- ☐ None or very few (0–10 books)
- ☐ Around one shelf (11–25 books)
- ☐ Around one bookcase (26–100 books)
- ☐ Around two bookcases (101–200 books)
- ☐ Around five bookcases (201–500 books)
- ☐ More than five bookcases (more than 500 books)

What is the highest level of schooling completed by your mother and father? (Or your step-parents or other guardians.) Don't worry if you're not sure, or if you don't know; just answer as closely as you can, or leave the questions blank.

Mother	Father	
☐	☐	GCSE or equivalent qualifications at secondary school
☐	☐	A-Level or equivalent qualifications at secondary school or college
☐	☐	Vocational qualifications at secondary school or college
☐	☐	Vocational qualifications after secondary school or college
☐	☐	A first degree at university (Bachelors)
☐	☐	A further degree at university (Masters or a Doctorate)

Do either of your parents (or other guardians) work in any job or area related to science?

☐ Yes	☐ No

Puzzle 1

Which of the following defines a compound?

- ☐ Different substances mixed together
- ☐ Atoms and molecules mixed together
- ☐ Atoms of different elements combined together
- ☐ Atoms of the same element combined together

	Not at all confident	Not very confident	Confident	Very confident
How confident are you that you solved this correctly?	☐	☐	☐	☐

How are you are doing in science?

To what extent do you agree or disagree with these statements?	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
I usually do well in science.	☐	☐	☐	☐	☐	☐
I have always been good at science.	☐	☐	☐	☐	☐	☐
I get good grades in science.	☐	☐	☐	☐	☐	☐
I understand even the most difficult science work.	☐	☐	☐	☐	☐	☐
I learn things quickly in science.	☐	☐	☐	☐	☐	☐
I need to use much effort to do well in science.	☐	☐	☐	☐	☐	☐
Science is harder for me than any other subject.	☐	☐	☐	☐	☐	☐
Science is harder for me than for many of my classmates.	☐	☐	☐	☐	☐	☐
Science makes me confused and nervous.	☐	☐	☐	☐	☐	☐
I often worry that it will be difficult for me in science classes.	☐	☐	☐	☐	☐	☐
I get very tense when I have to do science work.	☐	☐	☐	☐	☐	☐
I feel helpless when doing a science problem.	☐	☐	☐	☐	☐	☐
I worry that I will get poor grades in science.	☐	☐	☐	☐	☐	☐

Do you have separate classes for different science subjects (biology, chemistry, and physics)?

☐ Yes ☐ No

Are you in a set or group for science or any science subject based on your performance?

☐ A top set ☐ A middle set ☐ A bottom set ☐ We don't have sets

Please write in answers to the following questions, using your grades, national curriculum levels, or other results.

What overall grade/level did you get last year for science?	
What overall grade/level have you got so far this year in science?	
On average, what grade/level do you usually get across all subjects?	
What is the lowest grade/level on your next science exam that you would be satisfied with getting?	
What grade/level do you think your class or set is generally getting in science?	
What grade/level do you think your friends are generally getting in science?	
What grade/level do you think you will be able to get at your next science exam (such as at the end of the year)?	

Please tick a grade for each of the following questions.	A*	A	B	C	D	E	Lower
What grade do you think you will be able to get at GCSE (or equivalent) science?	☐	☐	☐	☐	☐	☐	☐
What grade do you think you would be able to get if you studied your best science subject at A-Level?	☐	☐	☐	☐	☐	☐	☐
What grade do you think people need to get in order to be 'good' at science?	☐	☐	☐	☐	☐	☐	☐

When you think about how good you are at science, how important or influential are the following areas to you? (Put another way: could the following areas change your own view about how good you are at science? Would the area be unimportant and irrelevant to you, and would not cause you to change your view of how good you are? Or would the area be important and relevant, and would change your view of how good you are?)	Very unimportant	Unimportant	Slightly unimportant	Slightly important	Important	Very important
My ability to successfully complete tasks, exercises, or assignments.	☐	☐	☐	☐	☐	☐
The amount of effort I need to successfully complete tasks, exercises, or assignments.	☐	☐	☐	☐	☐	☐
My earlier grades, marks, and other results.	☐	☐	☐	☐	☐	☐
Comparing my abilities or grades in science against my abilities or grades in another subject.	☐	☐	☐	☐	☐	☐
Comparing my abilities or grades in science against other students.	☐	☐	☐	☐	☐	☐
Seeing other students being able to successfully complete tasks, exercises, or assignments.	☐	☐	☐	☐	☐	☐
Being told that I am good at science.	☐	☐	☐	☐	☐	☐
Being told that I am bad at science.	☐	☐	☐	☐	☐	☐
Being told that I can successfully complete tasks, exercises, or assignments in science.	☐	☐	☐	☐	☐	☐
Feeling that doing science work is interesting.	☐	☐	☐	☐	☐	☐
Enjoying doing science work.	☐	☐	☐	☐	☐	☐
Feeling anxious or stressed when I do science work.	☐	☐	☐	☐	☐	☐

Puzzle 2

Which of the following best describes the purpose of cellular respiration?

- ☐ To provide energy for cell activities.
- ☐ To produce sugar for storage in cells.
- ☐ To release oxygen for breathing.
- ☐ To supply carbon dioxide for photosynthesis.

	Not at all confident	Not very confident	Confident	Very confident
How confident are you that you solved this correctly?	☐	☐	☐	☐

How are you doing in other subjects?

To what extent do you agree or disagree with these statements?	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
I usually do well in English.	☐	☐	☐	☐	☐	☐
I usually do well in maths.	☐	☐	☐	☐	☐	☐
I usually do well in biology.	☐	☐	☐	☐	☐	☐
I usually do well in chemistry.	☐	☐	☐	☐	☐	☐
I usually do well in physics.	☐	☐	☐	☐	☐	☐
I usually do well in history.	☐	☐	☐	☐	☐	☐
I usually do well in geography.	☐	☐	☐	☐	☐	☐
I usually do well in modern foreign languages.	☐	☐	☐	☐	☐	☐
I usually do well in design and technology.	☐	☐	☐	☐	☐	☐
I usually do well in art and design.	☐	☐	☐	☐	☐	☐
I usually do well in music.	☐	☐	☐	☐	☐	☐
I usually do well in physical education.	☐	☐	☐	☐	☐	☐
I usually do well in citizenship.	☐	☐	☐	☐	☐	☐
I usually do well in information technology.	☐	☐	☐	☐	☐	☐

Puzzle 3

A parachute jumper is in different positions at different times during their jump.

- Position 1: In the aircraft before the jump.
- Position 2: In freefall immediately after jumping and before the parachute opens.
- Position 3: Falling to the ground after the parachute opens.
- Position 4: On the ground just after landing.

In which of the positions does the force of gravity act on the jumper?
☐ Position 2 only
☐ Positions 2 and 3 only
☐ Positions 1, 2, and 3 only
☐ Positions 1, 2, 3, and 4

	Not at all confident	Not very confident	Confident	Very confident
How confident are you that you solved this correctly?	☐	☐	☐	☐

Science and other people

To what extent do you agree or disagree with these statements?	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
Most of my friends do well in science.	☐	☐	☐	☐	☐	☐
Most of my friends work hard at science.	☐	☐	☐	☐	☐	☐
My friends enjoy taking science tests.	☐	☐	☐	☐	☐	☐
My parents believe it's important for me to study science.	☐	☐	☐	☐	☐	☐
My parents believe that science is important for my career.	☐	☐	☐	☐	☐	☐
My parents like science.	☐	☐	☐	☐	☐	☐
I have been praised for my ability in science.	☐	☐	☐	☐	☐	☐
I have been told that I am bad at science.	☐	☐	☐	☐	☐	☐
My science teacher tells me I am good at science.	☐	☐	☐	☐	☐	☐
My science teacher thinks I can do well in science work with difficult materials.	☐	☐	☐	☐	☐	☐
My science teacher tells me that I am bad at science.	☐	☐	☐	☐	☐	☐
When I see how another student solves a science problem, I can see myself solving the problem in the same way.	☐	☐	☐	☐	☐	☐
I get discouraged when I see other students do well in science.	☐	☐	☐	☐	☐	☐
I aim to understand and learn the material in science.	☐	☐	☐	☐	☐	☐
I aim to avoid looking like I don't understand the material in science.	☐	☐	☐	☐	☐	☐
I aim to perform better than other students in science.	☐	☐	☐	☐	☐	☐
I aim to avoid looking like I'm performing worse than other students in science.	☐	☐	☐	☐	☐	☐
My teacher aims to make us understand and learn the material in science.	☐	☐	☐	☐	☐	☐
My teacher aims to make us compete and perform against one another in science.	☐	☐	☐	☐	☐	☐

Puzzle 4

Please complete the table below to show the number of atoms of each element in a molecule of sulphuric acid (H_2SO_4).

Element	Hydrogen	Sulphur	Oxygen
Number of atoms			

	Not at all confident	Not very confident	Confident	Very confident
How confident are you that you solved this correctly?	☐	☐	☐	☐

More about other people

How often do you compare your abilities and grades in science:	Never	Rarely	Some-times	Often	Always or almost always
Against your abilities and grades in other subjects?	☐	☐	☐	☐	☐
Against people in your class?	☐	☐	☐	☐	☐
Against people in your school (outside of your class)?	☐	☐	☐	☐	☐
Against people in other schools?	☐	☐	☐	☐	☐

Thinking about science, to what extent do you agree or disagree with these statements?	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
I compare myself against people doing better than me.	☐	☐	☐	☐	☐	☐
I compare myself against people doing about the same as me.	☐	☐	☐	☐	☐	☐
I compare myself against people doing worse than me.	☐	☐	☐	☐	☐	☐
I don’t tend to compare myself against other people.	☐	☐	☐	☐	☐	☐
When someone is good in science, it means that they must be worse in some other subjects like English.	☐	☐	☐	☐	☐	☐
I know how well everyone in my class performs.	☐	☐	☐	☐	☐	☐
I know how well everyone in my school year performs.	☐	☐	☐	☐	☐	☐
I know how well my school performs on average compared to other schools.	☐	☐	☐	☐	☐	☐

Puzzle 5

Some birds eat snails. A species of snail that lives in the forest has a dark shell. The same species of snail that lives in a field has a light-coloured shell. **Please explain how this difference in shell colours helps the snails to survive.**

	Not at all confident	Not very confident	Confident	Very confident
How confident are you that you solved this correctly?	☐	☐	☐	☐

About science

To what extent do you agree or disagree with these statements?	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
I look forward to my science lessons.	☐	☐	☐	☐	☐	☐
I am interested in the things I learn in science.	☐	☐	☐	☐	☐	☐
Science is important to me.	☐	☐	☐	☐	☐	☐
I enjoy learning science.	☐	☐	☐	☐	☐	☐
I wish I did not have to study science.	☐	☐	☐	☐	☐	☐
Science is boring.	☐	☐	☐	☐	☐	☐
I like science.	☐	☐	☐	☐	☐	☐
I like biology.	☐	☐	☐	☐	☐	☐
I like chemistry.	☐	☐	☐	☐	☐	☐
I like physics.	☐	☐	☐	☐	☐	☐
Making an effort in science is worth it because it will help me in the work that I want to do later on.	☐	☐	☐	☐	☐	☐
Learning science is worthwhile for me because it will improve my career prospects.	☐	☐	☐	☐	☐	☐
Science is an important subject for me because I need it for what I want to study later on.	☐	☐	☐	☐	☐	☐
I will learn many things in science that will help me get a job.	☐	☐	☐	☐	☐	☐
I need to do well in science to get the job I want.	☐	☐	☐	☐	☐	☐
I need to do well in science to get into the university of my choice.	☐	☐	☐	☐	☐	☐
I think learning science will help me in my daily life.	☐	☐	☐	☐	☐	☐
I need science to learn other school subjects.	☐	☐	☐	☐	☐	☐
I would like a job that involves using science.	☐	☐	☐	☐	☐	☐
It is important to do well in science.	☐	☐	☐	☐	☐	☐
Science is important to me personally.	☐	☐	☐	☐	☐	☐
Thinking scientifically is an important part of who I am.	☐	☐	☐	☐	☐	☐
Being able to do science helps me show other people who I am.	☐	☐	☐	☐	☐	☐
I have to give up a lot to do well in science.	☐	☐	☐	☐	☐	☐
Success in science means that I need to give up other activities I enjoy.	☐	☐	☐	☐	☐	☐
I have to sacrifice a lot of free time to be good at science.	☐	☐	☐	☐	☐	☐
I have to invest a lot of time to get good grades in science.	☐	☐	☐	☐	☐	☐

How important are these strategies for learning science?	Very unimportant	Unimportant	Slightly unimportant	Slightly important	Important	Very important
Thinking about what I know or don’t know, how I learn, and changing my own studying approaches to fit the course.	☐	☐	☐	☐	☐	☐
Keeping trying and putting in effort, even if I get low or high grades.	☐	☐	☐	☐	☐	☐
Organising and planning how and what to study.	☐	☐	☐	☐	☐	☐
Memorising facts and how to do things.	☐	☐	☐	☐	☐	☐
Linking ideas together, finding new insights or ways to do things.	☐	☐	☐	☐	☐	☐

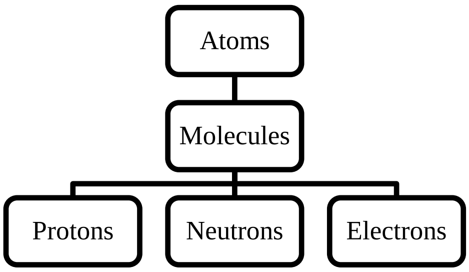
If you want, you can say more about what you think about science or what you like or dislike about any part of science.

Puzzle 6

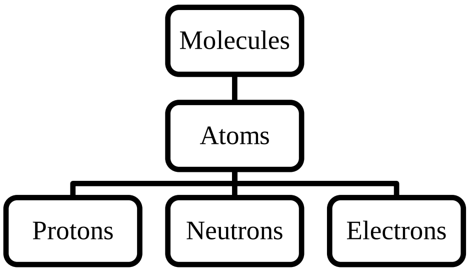
Which of these diagrams best represents the structure of matter, starting with the more complex particles at the top and ending with the more fundamental particles at the bottom?

☐ A ☐ B ☐ C ☐ D

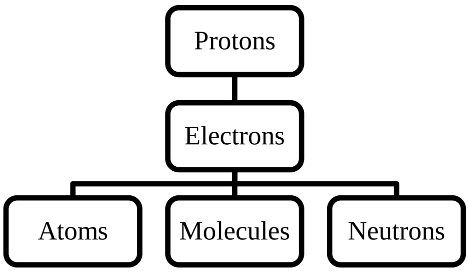
A



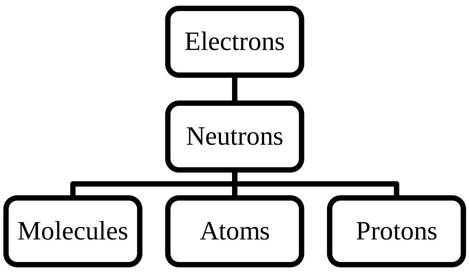
B



C



D



	Not at all confident	Not very confident	Confident	Very confident
How confident are you that you solved this correctly?	☐	☐	☐	☐

About ability in science

To what extent do you agree or disagree with these statements?	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
People can generally improve their ability in science.	☐	☐	☐	☐	☐	☐
I can improve my ability in science.	☐	☐	☐	☐	☐	☐
If I put in enough effort I can succeed in science.	☐	☐	☐	☐	☐	☐
Whether or not I do well in science is completely up to me.	☐	☐	☐	☐	☐	☐
If I wanted to, I could do well in science.	☐	☐	☐	☐	☐	☐
I do badly in science whether or not I study for my exams.	☐	☐	☐	☐	☐	☐
No matter whether or not I do my best in science, it will not improve my grades.	☐	☐	☐	☐	☐	☐
It is not worth practicing science for a test because I will come off badly again anyway.	☐	☐	☐	☐	☐	☐
I accomplish almost nothing in science of what I intend to do.	☐	☐	☐	☐	☐	☐

Puzzle 7

The amount of carbon dioxide in the air is increasing in a large city due to the growing number of vehicles. The mayor wants to plant more trees.

Do you agree with the mayor's suggestion?

☐ Yes ☐ No

Please explain why you agree or not.

	Not at all confident	Not very confident	Confident	Very confident
How confident are you that you solved this correctly?	☐	☐	☐	☐

About science classes

To what extent do you agree or disagree with these statements?	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
I know what my science teacher expects me to do.	☐	☐	☐	☐	☐	☐
I think of things not related to the science lesson.	☐	☐	☐	☐	☐	☐
My science teacher is easy to understand.	☐	☐	☐	☐	☐	☐
I am interested in what my science teacher says.	☐	☐	☐	☐	☐	☐
My science teacher gives me interesting things to do.	☐	☐	☐	☐	☐	☐
My science teacher gives me feedback on my strengths and weaknesses in science.	☐	☐	☐	☐	☐	☐
My science teacher tells us what is expected of us when we get a test, quiz or assignment.	☐	☐	☐	☐	☐	☐
My science teacher tells us what we have to learn.	☐	☐	☐	☐	☐	☐
My science teacher tells me what I need to do to become better in science.	☐	☐	☐	☐	☐	☐
My science teacher tells me about careers and jobs in science.	☐	☐	☐	☐	☐	☐
Other people at my school tell me about careers and jobs in science.	☐	☐	☐	☐	☐	☐
My school does special activities, talks, events, or visits	☐	☐	☐	☐	☐	☐

To what extent do you agree or disagree with these statements?	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
related to science (inside or outside of lessons).						

If you want, you can say more about your lessons or teachers.

Puzzle 8

As a liquid changes into a gas, which characteristics or properties change and which stay the same? In each row of the table below, please put a tick in the appropriate column.

	Changes	**Stays the same**
Density		
Mass		
Volume		
Size of molecules		
Speed of molecules		

	Not at all confident	Not very confident	Confident	Very confident
How confident are you that you solved this correctly?	☐	☐	☐	☐

About the future

How far in your education do you expect to go? (Please tick as many items as you need.)

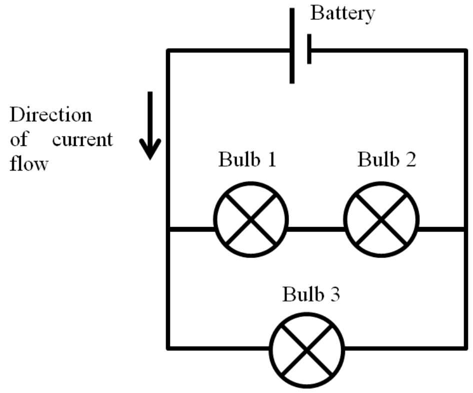
☐ I intend to study at A-Level or equivalent qualifications at secondary school or college
☐ I intend to study vocational qualifications at secondary school or college
☐ I intend to study vocational qualifications after secondary school or college
☐ I intend to study a first degree at university (Bachelors)
☐ I intend to study a further degree at university (Masters or a Doctorate)

To what extent do you agree or disagree with these statements?	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
I have thought about what I will study at GCSE (or equivalent).	☐	☐	☐	☐	☐	☐
I have thought about what I will study at A-Level (or equivalent).	☐	☐	☐	☐	☐	☐
I intend to study science at A-Level (or equivalent).	☐	☐	☐	☐	☐	☐
I have thought about what I will study at university.	☐	☐	☐	☐	☐	☐
I intend to study science at university.	☐	☐	☐	☐	☐	☐
I am planning on pursuing a career that involves a lot of science.	☐	☐	☐	☐	☐	☐

What subjects are you planning to study at A-Level (or equivalent)?

Puzzle 9

Three identical light bulbs are connected to a battery as shown in the diagram. The arrow indicates the direction of the current flow.



Which statement is true?

☐ The current in Bulb 1 is greater than the current in Bulb 2.
☐ The current in Bulb 1 is greater than the current in Bulb 3.
☐ The current in Bulb 2 is the same as the current in Bulb 3.
☐ The current in Bulb 2 is the same as the current in Bulb 1.

	Not at all confident	Not very confident	Confident	Very confident
How confident are you that you solved this correctly?	☐	☐	☐	☐

About your studying

To what extent do you agree or disagree with these statements?	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
During science class time I often miss important points because I'm thinking of other things.	☐	☐	☐	☐	☐	☐
When studying for science, I make up questions to help focus my studying.	☐	☐	☐	☐	☐	☐
When I become confused about something I'm studying for science class, I go back and try to figure it out.	☐	☐	☐	☐	☐	☐
If science course materials are difficult to understand, I change the way I approach the material.	☐	☐	☐	☐	☐	☐
Before I study new science course material thoroughly, I often skim it to see how it is organized.	☐	☐	☐	☐	☐	☐
I ask myself questions to make sure I understand the material I have been studying earlier in science class.	☐	☐	☐	☐	☐	☐
I try to change the way I study in order to fit the science course requirements and the teacher's teaching style.	☐	☐	☐	☐	☐	☐
I often find that I have been studying for science class but don't know what it was all about.	☐	☐	☐	☐	☐	☐
I try to think through a science topic and decide what I am supposed to learn from it rather than just reading it over when studying for science.	☐	☐	☐	☐	☐	☐
When studying for this science course I try to determine which concepts I don't understand well.	☐	☐	☐	☐	☐	☐
When I study for my science class, I set goals for myself in order to direct my activities in each study period.	☐	☐	☐	☐	☐	☐
If I get confused taking notes in science class, I make sure I sort it out afterwards.	☐	☐	☐	☐	☐	☐
When I study for a science test, I try to work out what the	☐	☐	☐	☐	☐	☐

To what extent do you agree or disagree with these statements?	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
most important parts to learn are.						
When I study science, I try to figure out which concepts I still have not understood properly.	☐	☐	☐	☐	☐	☐
When I study science, I start by working out exactly what I need to learn.	☐	☐	☐	☐	☐	☐
When I cannot understand something in science, I always search for more information to clarify the problem.	☐	☐	☐	☐	☐	☐
When I study for a science test, I learn as much as I can off by heart.	☐	☐	☐	☐	☐	☐
When I study science, I go over some problems so often that I feel as if I could solve them in my sleep.	☐	☐	☐	☐	☐	☐
In order to remember the method for solving a science problem, I go through examples again and again.	☐	☐	☐	☐	☐	☐
When I study science, I make myself check to see if I remember the work I have already done.	☐	☐	☐	☐	☐	☐
When I study for a science test, I try to understand new concepts by relating them to things I already know.	☐	☐	☐	☐	☐	☐
When I study science, I think of new ways to get the answer.	☐	☐	☐	☐	☐	☐
When I study science, I try to relate the work to things I have learnt in other subjects.	☐	☐	☐	☐	☐	☐
I think about how the science I have learnt can be used in everyday life.	☐	☐	☐	☐	☐	☐

Puzzle 10

An ice block is put in a bowl. Another ice block is wrapped in newspaper, and then put in another bowl.

Which ice block will melt first?

☐ The ice block not wrapped in newspaper
☐ The ice block wrapped in newspaper

Please explain your answer.

	Not at all confident	Not very confident	Confident	Very confident
How confident are you that you solved this correctly?	☐	☐	☐	☐

Final thoughts

Do you have any other comments about science at school?

Thank you for your help!

Appendix 4: England 2014/2015 survey: coding current grades/levels

Table A4.1: England 2014/2015 (Years 9, 10, 11): coding current grades/levels

National Curriculum Level	Year 7	Year 8	Year 9
8a and above	6	6	6
8b	6	6	6
8c	6	6	6
7a	6	6	5
7b	6	6	5
7c	6	5	5
6a	6	5	4
6b	5	5	4
6c	5	4	4
5a	5	4	3
5b	4	4	3
5c	4	3	3
4a	4	3	2
4b	3	3	2
4c	3	2	2
3a	3	2	1
3b	2	2	1
3c	2	1	1
2a	2	1	1
2b	1	1	1
2c and below	1	1	1

Note: Alphabetical grades were coded as: (6) A, (5) A, (4) B, (3) C, (2), D, (1) E and lower. The relative per year National Curriculum codes scaled as: (6) above an expected level [A*], (5) above an expected level [A], (4) at an expected level [B], (3) just below an expected level [C], (2), below an expected level [D], (1) below the expected level [E and lower].*

Appendix 5: England 2014/2015 survey: task-score details

Table A5.1: England 2014/2015 (Years 9, 10, 11): task-score reliability

		Reliability			
Factor/scale	Items	Year 9	Year 10	Year 11	Years 9-11
Task-score (10 items)	10	.661	.611	.631	.645
Task-confidence (10 items)	10	.904	.883	.897	.897
Task-score (8 items)	8	.617	.577	.595	.598
Task-confidence (8 items)	8	.882	.863	.875	.874

Notes: Reliability was measured through Cronbach's alpha (α) coefficients for the available students, depending on the questionnaire version.

Table A5.2: England 2014/2015 (Years 9, 10, 11): task-scores and task-confidence, item-level proportion correct

	Science task		TIMSS 2011 (Year 9) England results		2014/2015 survey (Year 9) results		2014/2015 survey (Year 10) results		2014/2015 survey (Year 11) results	
2014/2015 task number	TIMSS task number	Domain	M	SD	M	SD	M	SD	M	SD
1	S042306	Chemistry	.49	.50	.53	.50	.80	.40	.64	.48
2	S032611	Biology	.36	.48	.55	.50	.59	.49	.61	.49
3*	S032141	Physics	.44	.50	.54	.50	.74	.44	.55	.50
4	S042076	Chemistry	.46	.50	.58	.49	.75	.43	.71	.45
5*	S032451	Biology	.82	.37	.83	.34	.89	.28	.90	.26
6	S032579	Chemistry	.34	.47	.45	.50	.59	.49	.50	.50
7	S052091	Biology	.52	.50	.56	.50	.68	.47	.70	.46
8	S042173Z	Physics	.65	.30	.71	.31	.81	.28	.79	.27
9	S032184	Physics	.46	.50	.48	.50	.48	.50	.48	.50
10	S042407	Physics	.38	.49	.21	.41	.16	.37	.11	.32

Notes: Means (M) and standard deviations (SD) are reported. Missing responses were not estimated. Partial-credit tasks were coded as 0-1 proportions (0 for incorrect, .5 for partially correct, and 1.0 for fully correct). The means reflect proportions correct for those who were administered the item and who attempted the item. In the 2014/2015 survey, items 3 and 5 were not included in the last questionnaire version (i.e. the proportions are only reported for those who were administered the items).

Appendix 6: England 2014/2015 survey: calculating sensitivity, specificity, and simple-matching

Table A6.1: England 2014/2015 (Years 9, 10, 11): task-score and task-confidence intersections

	Task-score	
Task-confidence	Correct answer	Incorrect answer
Confident	True positive (A)	False positive (B)
Not confident	False negative (C)	True negative (D)

Indicators of sensitivity, specificity, and simple-matching were calculated from the intersection of task-score and task-confidence. The various categories were summed for each student across their respective responses; the following formulae were then used (Schraw, Kuch, & Gutierrez, 2013; Schraw, Kuch, Gutierrez, & Richmond, 2014; Yule, 1912).

$$\text{Task sensitivity} = A / (A + C)$$
$$\text{Task specificity} = D / (B + D)$$
$$\text{Task simple-matching} = (A + D) / (A + B + C + D)$$

Task sensitivity reflected the number of 'true positive' ratings across all the items, as a proportion of the total number of 'true positive' and 'false negative' ratings. Sensitivity provided an indicator of when someone knows that they have answered correctly (i.e. being 'confident' when they have the right answer).

Task specificity reflected the number of 'true negative' ratings across all the items, as a proportion of the total number of 'true negative' and 'false positive' ratings. Specificity provided an indicator of when someone knows that they have the wrong answer (i.e. being 'not confident' when they have the wrong answer).

Task simple-matching reflected the number of 'true positive' and 'true negative' ratings across all the items, as a proportion of all ratings. Simple-matching provided an indicator of the combined proportions of both knowing when answers are right and knowing when answers are wrong.

Within statistics and research methods, cell B reflects a Type I error (a false positive) and cell C reflects a Type II error (a false negative). Specificity subtracted from one represents a Type I (false positive) error rate (which can also be calculated directly as B / (B+D)), and sensitivity subtracted from one represents a Type II (false negative) error rate (which can also be calculated directly as C / (A + C)).

Appendix 7: England 2014/2015 survey: responses per academic year

Table A7.1: England 2014/2015 (Years 9, 10, 11): descriptive statistics per year

	Year 9 (A)		Year 10 (B)		Year 11 (C)	
Item/factor (1-6 scales unless highlighted otherwise)	M	SD	M	SD	M	SD
Science intentions	[AC] 3.62	1.42	[BC] 3.57	1.43	[AC BC] 3.08	1.57
Task-score (0-1)	[AB AC] .50	.28	[AB] .62	.28	[AC] .58	.29
Task-confidence (0-1)	[AB] .52	.22	[AB BC] .57	.25	[BC] .51	.23
Task-confidence accuracy/bias (-1 to +1)	[AB AC] .02	.25	[AB] -.05	.26	[AC] -.06	.26
Task sensitivity (0-1)	[AB] .59	.35	[AB BC] .67	.33	[BC] .59	.33
Task specificity (0-1)	.57	.36	.56	.37	.62	.36
Task simple-matching (0-1)	[AB] .62	.22	[AB] .69	.24	.66	.23
Gender (1=boy)	[AB] .55	.50	[AB BC] .67	.47	[BC] .50	.50
Ethnicity (1=White)	[(ALL)] .64	.48	[(ALL)] .55	.50	[(ALL)] .83	.37
Ethnicity (1=Black)	[AB] .02	.13	[AB] .06	.23	.04	.19
Ethnicity (1=East-Asian)	.02	.13	.03	.17	.01	.12
Ethnicity (1=South-Asian/Indian)	[AC] .27	.45	[BC] .29	.46	[AC BC] .07	.25
Ethnicity (1=mixed)	.05	.21	.05	.21	.04	.20
Ethnicity (1=other)	.01	.09	.02	.14	.01	.08
Books at home (1-5)	3.12	1.30	3.27	1.33	3.13	1.32
Parent(s) in science (1=yes)	.22	.42	[BC] .25	.43	[BC] .16	.37
Mothers' education	[AC] 3.05	1.55	[BC] 3.19	1.66	[AC BC] 2.54	1.60
Fathers' education	[(ALL)] 3.12	1.59	[(ALL)] 3.38	1.74	[(ALL)] 2.74	1.66
Current grade	[AB] 3.25	1.66	[AB BC] 3.93	1.42	[BC] 3.30	1.50
Self-concept	[AC] 3.93	1.04	[BC] 3.81	1.12	[AC BC] 3.62	1.10
Self-efficacy	[AC] 4.57	1.09	[BC] 4.68	1.09	[AC BC] 3.81	1.18
Interest value	[AC] 4.10	1.25	[BC] 4.09	1.14	[AC BC] 3.79	1.22
Utility value	[(ALL)] 4.34	1.16	[(ALL)] 4.15	1.12	[(ALL)] 3.83	1.18
Personal value	[AC] 3.66	1.40	[BC] 3.57	1.37	[AC BC] 3.25	1.33

Item/factor (1-6 scales unless highlighted otherwise)	Year 9 (A)		Year 10 (B)		Year 11 (C)	
	M	SD	M	SD	M	SD
Cost value (absence of)	[AB] 3.84	1.34	[AB] 3.57	1.22	3.70	1.15
Orientation: mastery	[AC] 4.71	1.11	[BC] 4.73	1.15	[AC BC] 4.47	1.12
Orientation: performance	[AC] 4.45	1.31	[BC] 4.40	1.26	[AC BC] 4.01	1.29
Perceived control	[AB AC] 4.77	.89	[AB] 4.58	1.02	[AC] 4.48	.94
Perceived control (exams)	[AB AC] 4.32	1.26	[AB] 4.15	1.23	[AC] 4.08	1.17
Study strategy: self-regulation	[AC] 3.76	.80	[BC] 3.75	.68	[AC BC] 3.62	.75
Study strategy: control	[AC] 4.21	1.05	4.13	.92	[AC] 4.00	1.01
Study strategy: memorisation	3.98	1.03	3.98	.97	3.83	.99
Study strategy: elaboration	[AC] 3.96	1.08	3.82	1.01	[AC] 3.70	.99
Anxiety (absence of)	4.19	1.25	4.22	1.23	4.09	1.19
Mastery norms (good grade)	(ALL) 4.35	.99	(ALL) 4.52	.98	(ALL) 4.11	.96
Subject-comparisons	[AC] 4.07	1.50	[BC] 4.01	1.57	[AC BC] 3.73	1.51
Peer-comparisons	[AC] 4.20	1.32	4.14	1.38	[AC] 3.96	1.33
Social persuasions (praise)	3.83	1.20	3.88	1.14	3.78	1.10
Vicarious experiences	[AC] 4.09	1.29	[BC] 4.02	1.22	[AC BC] 3.71	1.16
Norms/influence (friends)	[AB] 3.81	.86	[AB] 3.65	.97	3.69	.86
Norms/influence (parents)	(ALL) 4.44	1.16	(ALL) 4.25	1.17	(ALL) 3.83	1.10
Teacher perceptions	4.34	1.01	4.32	.96	4.26	.91
Teacher/school careers/events	3.55	1.19	3.63	1.10	3.50	1.07

Notes: Missing responses were estimated by expectation-maximisation. Means (M) and standard deviations (SD) are reported. Differences in means across academic years ($p < .05$) are highlighted in superscript; for brevity, 'ALL' indicates where all years differed. Books at home were coded as in international surveys: (1) 0-10 books; (2) 11-25 books; (3) 26-100 books; (4) 101-200 books; and (5) more than 200 books. Parental education was coded as: (1) GCSE or equivalent qualifications at secondary school; (2) A-Level or equivalent qualifications at secondary school or college; (3) vocational qualifications at secondary school or college; (4) vocational qualifications after secondary school or college; (5) a first degree at university (Bachelors); and (6) a further degree at university (Masters or a Doctorate).

Appendix 8: England 2014/2015 survey: correlation tables

Table A8.1: England 2014/2015 (Years 9, 10, 11): correlations (set 1)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Science intentions	1.000												
2. Current grade	.351	1.000											
3. Self-concept	.475	.491	1.000										
4. Self-efficacy	.444	.638	.553	1.000									
5. Interest value	.622	.424	.642	.484	1.000								
6. Utility value	.732	.329	.505	.422	.724	1.000							
7. Personal value	.684	.359	.525	.416	.723	.777	1.000						
8. Cost value (absence of)	-.135	-.053	(.044)	(.016)	-.080	-.216	-.250	1.000					
9. Anxiety (absence of)	.309	.362	.549	.416	.461	.255	.307	.246	1.000				
10. Norms (parents)	.554	.297	.393	.420	.527	.668	.543	-.149	.221	1.000			
11. Teacher perceptions	.399	.249	.435	.252	.628	.507	.484	-.066	.318	.353	1.000		
12. Careers/events	.374	.263	.284	.194	.411	.376	.449	-.234	.138	.309	.548	1.000	
13. Gender	.152	.241	.191	.279	.190	.107	.195	(.016)	.255	.120	.083	.080	1.000
14. Year	-.132	(.049)	-.111	-.221	-.090	-.169	-.111	-.058	(-.029)	-.197	(-.029)	(-.010)	(-.013)

Notes: Missing responses were estimated by expectation-maximisation. Pearson's correlation coefficients (R) are reported. Coefficients in brackets highlight non-significant associations (i.e. not meeting the $p < .05$ criterion for significance).

Table A8.2: England 2014/2015 (Years 9, 10, 11): correlations (set 2)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Science intentions	1.000												
2. Current grade	.351	1.000											
3. Self-concept	.475	.491	1.000										
4. Self-efficacy	.444	.638	.553	1.000									
5. Task-score	.256	.517	.328	.422	1.000								
6. Task-confidence	.403	.514	.560	.531	.520	1.000							
7. Task-conf. accuracy/bias (diff.-s.)	.080	-.108	.141	(.012)	-.638	.326	1.000						
8. Task sensitivity	.259	.389	.424	.401	.374	.744	.256	1.000					
9. Task specificity	-.317	-.238	-.410	-.332	-.171	-.679	-.423	-.403	1.000				
10. Task simple-matching	.115	.273	.144	.214	.281	.233	-.101	.561	.255	1.000			
11. Task-conf. accuracy/bias (resid.)	.298	.285	.448	.346	(.000)	.849	.769	.643	-.687	.109	1.000		
12. Self-concept accuracy/bias (resid.)	.337	(.000)	.845	.267	.132	.363	.203	.301	-.325	.069	.358	1.000	
13. Gender	.152	.241	.191	.279	.187	.348	.107	.296	-.248	.160	.286	.130	1.000
14. Year	-.132	(.049)	-.111	-.221	.134	(.019)	-.131	(.028)	(.044)	.071	(.000)	(.000)	-.013

Notes: Missing responses were estimated by expectation-maximisation. Pearson's correlation coefficients (R) are reported. Coefficients in brackets highlight non-significant associations (i.e. not meeting the p < .05 criterion for significance).

Table A8.3: England 2014/2015 (Years 9, 10, 11): correlations (set 3)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Science intentions	1.000												
2. Current grade	.351	1.000											
3. Self-concept	.475	.491	1.000										
4. Self-efficacy	.444	.638	.553	1.000									
5. Norms/influence (friends)	.162	.119	.189	.110	1.000								
6. Norms/influence (parents)	.554	.297	.393	.420	.316	1.000							
7. Teacher perceptions	.399	.249	.435	.252	.299	.353	1.000						
8. Teacher/school careers/events	.374	.263	.284	.194	.216	.309	.548	1.000					
9. Orientation: mastery	.357	.316	.366	.365	.269	.468	.372	.187	1.000				
10. Orientation: performance	.317	.249	.359	.368	.150	.429	.276	.145	.462	1.000			
11. Perceived control	.361	.263	.450	.374	.290	.413	.547	.208	.438	.370	1.000		
12. Perceived control (exams)	.190	.270	.446	.358	(.035)	.221	.269	(-.018)	.251	.204	.302	1.000	
13. Gender	.152	.241	.191	.279	(-.041)	.120	.083	.080	(.036)	.154	.060	.132	1.000
14. Year	-.132	(.049)	-.111	-.221	-.065	-.197	(-.029)	(-.010)	-.072	-.125	-.123	-.083	(-.013)

Notes: Missing responses were estimated by expectation-maximisation. Pearson's correlation coefficients (R) are reported. Coefficients in brackets highlight non-significant associations (i.e. not meeting the p < .05 criterion for significance).

Table A8.4: England 2014/2015 (Years 9, 10, 11): correlations (set 4)

	1	2	3	4	5	6	7	8	9
1. Science intentions	1.000								
2. Current grade	.351	1.000							
3. Self-concept	.475	.491	1.000						
4. Self-efficacy	.444	.638	.553	1.000					
5. Study strategy: self-regulation	.492	.365	.519	.369	1.000				
6. Study strategy: control	.444	.353	.441	.340	.745	1.000			
7. Study strategy: memorisation	.435	.305	.365	.309	.711	.700	1.000		
8. Study strategy: elaboration	.448	.266	.420	.298	.671	.610	.664	1.000	
9. Gender	.152	.241	.191	.279	.130	(.016)	.057	.115	1.000
10. Year	-.132	(.049)	-.111	-.221	-.066	-.081	-.051	-.102	(-.013)

Notes: Missing responses were estimated by expectation-maximisation. Pearson's correlation coefficients (R) are reported. Coefficients in brackets highlight non-significant associations (i.e. not meeting the p < .05 criterion for significance).

Table A8.5: England 2014/2015 (Years 9, 10, 11): correlations (set 5)

	1	2	3	4	5	6	7	8	9	10	11
1. Science intentions	1.000										
2. Current grade	.351	1.000									
3. Self-concept	.475	.491	1.000								
4. Self-efficacy	.444	.638	.553	1.000							
5. Anxiety (absence of)	.309	.362	.549	.416	1.000						
6. Mastery norms (grade)	.122	.288	.066	.312	(.037)	1.000					
7. Subject-comparisons	.331	.316	.498	.395	.672	(.028)	1.000				
8. Peer-comparisons	.261	.306	.497	.380	.672	(.043)	.632	1.000			
9. Social persuasions (praise)	.389	.378	.600	.414	.395	.052	.348	.358	1.000		
10. Vicarious experiences	.240	.229	.363	.314	.238	.063	.206	.220	.429	1.000	
11. Gender	.152	.241	.191	.279	.255	.144	.215	.162	.146	.115	1.000
12. Year	-.132	(.049)	-.111	-.221	(-.029)	-.068	-.084	-.068	(-.011)	-.112	(-.013)

Notes: Missing responses were estimated by expectation-maximisation. Pearson's correlation coefficients (R) are reported. Coefficients in brackets highlight non-significant associations (i.e. not meeting the $p < .05$ criterion for significance).

Table A8.6: England 2014/2015 (Years 9, 10, 11): correlations with science intentions, summary

	Correlation with science intentions			
Item/factor	Year 9	Year 10	Year 11	Years 9-11
Year	NA	NA	NA	-.132
Gender (1=boy)	.077	.204	.194	.152
Task-score	.208	.329	.313	.256
Task-confidence	.333	.470	.444	.403
Task-confidence accuracy/bias (difference-score)	(.064)	.094	(.037)	.080
Task sensitivity	.182	.296	.360	.259
Task specificity	-.277	-.325	-.362	-.317
Task simple-matching	(.061)	.163	.167	.115
Task-confidence accuracy/bias (regression-residual)	.263	.338	.327	.298
Self-concept accuracy/bias (regression-residual)	.346	.243	.458	.337
Ethnicity (White)	-.329	-.200	(-.067)	-.251
Ethnicity (Black)	(.025)	(.004)	(.004)	(.007)
Ethnicity (East-Asian)	(-.015)	(.041)	(.067)	(.026)
Ethnicity (South-Asian/Indian)	.327	.160	(.067)	.243
Ethnicity (mixed)	(.042)	(.019)	(-.008)	(.025)
Ethnicity (other)	(.023)	.109	(.014)	.058
Books at home	.093	.188	.121	.130
Parent(s) in science (1=yes)	(.044)	.122	.208	.115
Mothers' education	.142	.122	.219	.172
Fathers' education	.204	.185	.274	.227
Current grade	.326	.359	.416	.351
Self-concept	.452	.377	.613	.475
Self-efficacy	.369	.433	.509	.444
Interest value	.571	.666	.643	.622
Utility value	.712	.730	.751	.732
Personal value	.616	.740	.723	.684
Cost value (absence of)	-.174	-.143	(-.066)	-.135

Item/factor	Correlation with science intentions			
	Year 9	Year 10	Year 11	Years 9-11
Orientation: mastery	.347	.340	.366	.357
Orientation: performance	.275	.320	.334	.317
Perceived control	.354	.400	.291	.361
Perceived control (exams)	.142	.150	.309	.190
Study strategy: self-regulation	.450	.536	.509	.492
Study strategy: control	.448	.444	.419	.444
Study strategy: memorisation	.427	.437	.439	.435
Study strategy: elaboration	.422	.453	.472	.448
Anxiety (absence of)	.251	.325	.386	.309
Mastery norms (good grade)	.148	.132	(-.004)	.122
Subject-comparisons	.292	.346	.346	.331
Peer-comparisons	.210	.257	.327	.261
Social persuasions (praise)	.354	.410	.441	.389
Vicarious experiences	.181	.264	.270	.240
Norms/influence (friends)	.153	.198	.116	.162
Norms/influence (parents)	.534	.520	.592	.554
Teacher perceptions	.383	.461	.352	.399
Teacher/school careers/events	.360	.366	.418	.374

Notes: Missing responses were estimated by expectation-maximisation. Pearson's correlation coefficients (R) are reported. Coefficients in brackets highlight non-significant associations (i.e. not meeting the $p < .05$ criterion for significance).

Appendix 9: Further cluster solutions

Table A9.1: PISA (England) 2006: accuracy/bias clusters (five clusters)

	Cluster A		Cluster B		Cluster C		Cluster D		Cluster E	
Item/factor (z-scores)	M	SD	M	SD	M	SD	M	SD	M	SD
Self-concept accuracy/bias	-.35	.46	.65	.36	1.56	.63	-1.79	.67	.30	.31
Intentions	-.22	.85	.43	.89	.91	1.04	-1.02	.57	.12	.89
Gender (1=boy)	-.13	.99	.14	.99	.35	.94	-.28	.96	.14	.99
Books at home	-.05	.99	.13	1.00	.23	.98	-.21	1.01	.03	.95
Parent(s) in sci. (1=yes)	-.02	.98	.04	1.03	.15	1.10	-.17	.84	-.02	.99
Mothers' education	-.03	.99	.04	1.00	.18	1.06	-.13	1.04	-.02	.94
Fathers' education	-.06	.98	.10	.98	.28	1.09	-.22	1.07	-.01	.94
Task-score (PV1)	-.14	.88	.30	1.01	.60	1.08	-.53	.94	.14	.87
Self-concept	-.39	.36	.71	.00	1.69	.53	-1.90	.58	.32	.00
Self-efficacy (areas)	-.24	.80	.48	.95	.90	1.02	-.90	.96	.11	.73
Interest (various areas)	-.15	.87	.34	.87	.72	.91	-1.01	1.16	.18	.81
Interest value	-.24	.79	.43	.80	1.03	.98	-1.17	.86	.21	.78
Utility value	-.22	.84	.38	.90	.87	.94	-1.01	.95	.20	.87
Personal value	-.24	.82	.37	.84	.98	.99	-1.01	.90	.17	.84
General value	-.19	.87	.26	.94	.82	1.08	-.67	.94	.08	.88
Science activities	-.18	.92	.31	.94	.76	.98	-.71	.79	.07	.95
School career preparation	-.17	.89	.27	.94	.78	1.01	-.72	1.03	.09	.89
School career information	-.13	.93	.26	.94	.56	1.03	-.62	1.11	.06	.91
Teaching: interaction	-.09	.92	.13	.95	.42	1.12	-.55	1.16	.10	.93
Teaching: activities	-.09	.93	.13	.98	.36	1.07	-.45	1.15	.11	.98
Teaching: investigations	-.09	.96	.22	1.01	.28	1.12	-.43	.89	.00	.97
Teaching: applications	-.11	.92	.17	1.00	.52	1.11	-.59	1.05	.06	.90
Cluster size (N, %)	2390	50.5%	754	15.9%	646	13.6%	441	9.3%	502	10.6%

Notes: Missing responses were estimated by expectation-maximisation. Means (M) and standard deviations (SD) are reported. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey.*

Table A9.2: PISA (England) 2006: accuracy/bias clusters (five clusters), cluster differences

	Overall group difference		Paired group differences, sig. (*p*)									
Item/factor	Sig. (*p*)	Size (η^2)	A-B	A-C	A-D	A-E	B-C	B-D	B-E	C-D	C-E	D-E
Self-concept accuracy/bias	**<.001**	.767	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Intentions	**<.001**	.261	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Gender (1=boy)	**<.001**	.038	**<.001**	**<.001**	**.036**	**<.001**	**.001**	**<.001**	1.000	**<.001**	**.003**	**<.001**
Books at home	**<.001**	.015	**<.001**	**<.001**	**.022**	.901	.497	**<.001**	1.000	**<.001**	**.008**	**.002**
Parent(s) in sci. (1=yes)	**<.001**	.006	1.000	**.001**	**.042**	1.000	.489	**.004**	1.000	**<.001**	.055	.170
Mothers' education	**<.001**	.007	1.000	**<.001**	.423	1.000	.085	**.048**	1.000	**<.001**	**.010**	.771
Fathers' education	**<.001**	.018	**.001**	**<.001**	**.020**	1.000	.009	**<.001**	.655	**<.001**	**<.001**	**.010**
Task-score (PV1)	**<.001**	.103	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**.021**	**<.001**	**<.001**	**<.001**
Self-concept	**<.001**	.869	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-efficacy (areas)	**<.001**	.251	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Interest (various areas)	**<.001**	.196	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**.018**	**<.001**	**<.001**	**<.001**
Interest value	**<.001**	.330	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Utility value	**<.001**	.245	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**.004**	**<.001**	**<.001**	**<.001**
Personal value	**<.001**	.275	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
General value	**<.001**	.163	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**.012**	**<.001**	**<.001**	**<.001**
Science activities	**<.001**	.158	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
School career preparation	**<.001**	.157	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**.012**	**<.001**	**<.001**	**<.001**
School career information	**<.001**	.096	**<.001**	**<.001**	**<.001**	**.001**	**<.001**	**<.001**	**.003**	**<.001**	**<.001**	**<.001**
Teaching: interaction	**<.001**	.059	**<.001**	**<.001**	**<.001**	**.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**
Teaching: activities	**<.001**	.045	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**
Teaching: investigations	**<.001**	.039	**<.001**	**<.001**	**<.001**	.692	1.000	**<.001**	**.001**	**<.001**	**<.001**	**<.001**
Teaching: applications	**<.001**	.079	**<.001**	**<.001**	**<.001**	**.002**	**<.001**	**<.001**	.549	**<.001**	**<.001**	**<.001**

Notes: Missing responses were estimated by expectation-maximisation. Significant p-values (p < .05) are highlighted in bold. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey.*

Table A9.3: PISA (England) 2006: intention/attitude clusters (five clusters)

	Cluster A		Cluster B		Cluster C		Cluster D		Cluster E	
Item/factor (*z*-scores)	M	SD	M	SD	M	SD	M	SD	M	SD
Self-concept accuracy/bias	.39	.81	-.66	1.03	-.03	.74	-.33	.84	.86	.94
Intentions	.78	.56	-1.33	.00	.08	.00	-.60	.24	1.39	.77
Gender (1=boy)	.13	.99	-.17	.98	-.03	1.00	-.06	1.00	.03	1.00
Books at home	.13	.98	-.16	1.04	-.11	.97	.07	.97	.19	.98
Parent(s) in sci. (1=yes)	.04	1.03	-.06	.95	-.08	.93	-.03	.98	.19	1.12
Mothers' education	.08	1.02	-.05	1.02	-.09	.97	.02	.99	.09	1.05
Fathers' education	.14	.98	-.16	1.01	-.08	.99	-.03	.98	.20	1.06
Task-score (PV1)	.21	1.03	-.29	.89	-.14	.92	.04	.89	.56	.96
Self-concept	.43	.81	-.74	1.00	-.09	.70	-.31	.82	1.00	.93
Self-efficacy (areas)	.26	.94	-.43	1.01	-.09	.86	-.14	.86	.72	.96
Interest (various areas)	.38	.73	-.79	1.17	.04	.73	-.20	.69	.85	.84
Interest value	.48	.82	-.84	.87	-.06	.68	-.33	.73	1.10	.88
Utility value	.37	.61	-.91	.84	-.18	.62	-.40	.63	1.77	.00
Personal value	.46	.84	-.80	.86	-.14	.64	-.38	.63	1.27	.85
General value	.26	.99	-.44	.94	-.17	.82	-.21	.82	.88	1.00
Science activities	.41	.93	-.61	.84	-.11	.90	-.26	.86	.78	.94
School career preparation	.22	.91	-.48	1.02	-.11	.81	-.16	.89	.89	1.00
School career information	.25	.88	-.50	1.07	.02	.85	-.23	.94	.64	1.00
Teaching: interaction	.11	.96	-.30	1.06	.04	.89	-.06	1.00	.29	1.07
Teaching: activities	.13	.96	-.25	1.09	-.03	.97	-.06	.94	.29	.98
Teaching: investigations	.11	1.03	-.25	.95	.06	.98	-.13	.97	.16	1.05
Teaching: applications	.15	.96	-.35	1.06	-.04	.90	-.10	.94	.47	1.04
Cluster size (N, %)	1321	28.2%	1048	22.4%	989	21.1%	782	16.7%	537	11.5%

Notes: Missing responses were estimated by expectation-maximisation. Means (M) and standard deviations (SD) are reported. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey.*

Table A9.4: PISA (England) 2006: intention/attitude clusters (five clusters), cluster differences

Item/factor	Overall group difference: Sig. (p)	Overall group difference: Size (η^2)	Paired group differences, sig. (p): A-B	A-C	A-D	A-E	B-C	B-D	B-E	C-D	C-E	D-E
Self-concept accuracy/bias	**<.001**	.243	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Intentions	**<.001**	.837	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Gender (1=boy)	**<.001**	.012	**<.001**	**.001**	**<.001**	.409	**.015**	.180	**.002**	1.000	1.000	1.000
Books at home	**<.001**	.018	**<.001**	**<.001**	1.000	1.000	1.000	**<.001**	**<.001**	**.001**	**<.001**	.266
Parent(s) in sci. (1=yes)	**<.001**	.007	.101	**.033**	1.000	**.044**	1.000	1.000	**<.001**	1.000	**<.001**	**.001**
Mothers' education	**<.001**	.005	**.034**	**.001**	1.000	1.000	1.000	1.000	.103	.204	**.006**	1.000
Fathers' education	**<.001**	.017	**<.001**	**<.001**	**.002**	1.000	.765	.088	**<.001**	1.000	**<.001**	**<.001**
Task-score (PV1)	**<.001**	.072	**<.001**	**<.001**	**.001**	**<.001**	.005	**<.001**	**<.001**	**.001**	**<.001**	**<.001**
Self-concept	**<.001**	.299	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-efficacy (areas)	**<.001**	.126	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**
Interest (various areas)	**<.001**	.270	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Interest value	**<.001**	.374	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Utility value	**<.001**	.601	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Personal value	**<.001**	.408	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
General value	**<.001**	.161	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**
Science activities	**<.001**	.211	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**.003**	**<.001**	**<.001**
School career preparation	**<.001**	.161	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**
School career information	**<.001**	.127	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Teaching: interaction	**<.001**	.033	**<.001**	.855	**.001**	**.005**	**<.001**	**<.001**	**<.001**	.363	**<.001**	**<.001**
Teaching: activities	**<.001**	.028	**<.001**	**.001**	**<.001**	**.018**	**<.001**	**.001**	**<.001**	1.000	**<.001**	**<.001**
Teaching: investigations	**<.001**	.024	**<.001**	1.000	**<.001**	1.000	**<.001**	.089	**<.001**	**<.001**	.708	**<.001**
Teaching: applications	**<.001**	.059	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**

Notes: Missing responses were estimated by expectation-maximisation. Significant p-values ($p < .05$) are highlighted in bold. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey.*

Table A9.5: England 2014/2015 (Years 9, 10, 11): task-level accuracy/bias clusters (five clusters)

	Cluster A		Cluster B		Cluster C		Cluster D		Cluster E	
Item/factor (per year *z*-scores)	M	SD	M	SD	M	SD	M	SD	M	SD
Science intentions	.09	.96	-.33	.95	-.19	.87	.62	.85	1.00	.85
Task-score	.55	.46	-.71	.40	-1.90	.17	1.40	.05	1.18	.39
Task-confidence	.20	.73	-.59	.75	-.67	1.06	1.16	.61	1.80	.26
Task-conf. accuracy/bias (diff.-score)	-.43	.88	.25	.83	1.48	1.02	-.48	.57	.32	.52
Task sensitivity	.19	.81	-.48	1.10	-.41	.75	.86	.42	1.10	.14
Task specificity	-.06	1.04	.30	.84	.17	1.08	-.45	.85	-1.25	.54
Task simple-matching	-.06	.85	-.26	.88	-.12	1.63	1.00	.61	1.01	.42
Task-conf. accuracy/bias (regression-r.)	-.11	.93	-.26	.92	.38	1.28	.50	.72	1.39	.40
Self-concept accuracy/bias (regression-r.)	.03	.95	-.16	1.03	-.32	1.03	.22	.90	.84	.58
Gender (1=boy)	.04	.99	-.22	1.02	-.12	1.03	.39	.83	.78	.41
Ethnicity (White)	.01	.99	.24	.90	-.22	1.11	-.48	.97	-.60	.98
Ethnicity (Black)	-.02	.98	-.02	.94	.11	1.11	.07	1.17	.01	1.09
Ethnicity (East-Asian)	-.03	.89	-.08	.67	-.07	.80	.29	1.65	.51	2.10
Ethnicity (South-Asian/Indian)	.02	1.03	-.22	.78	.20	1.23	.29	1.05	.50	1.07
Ethnicity (mixed)	.01	1.02	-.04	.91	.09	1.17	.04	1.08	.00	1.04
Ethnicity (other)	-.06	.58	.01	1.12	-.01	1.01	.30	1.91	-.10	.02
Books at home	.23	.93	-.31	.97	-.67	1.01	.44	.81	.58	.81
Parent(s) in science (1=yes)	.02	1.01	-.13	.89	-.04	.98	.32	1.18	.21	1.13
Mothers' education	.11	1.00	-.33	.88	-.14	1.01	.57	.96	.76	.90
Fathers' education	.04	.99	-.32	.89	-.07	.97	.73	.89	.84	.85
Current grade	.23	.88	-.50	.85	-.63	.90	.93	.69	1.18	.64
Self-concept	.14	.87	-.38	.92	-.54	1.11	.64	.85	1.31	.60
Self-efficacy	.24	.85	-.51	.92	-.49	1.01	.86	.62	1.14	.38
Interest value	.17	.89	-.38	.99	-.49	.98	.62	.73	1.12	.63
Utility value	.11	.94	-.32	1.01	-.24	1.04	.56	.70	.88	.70
Personal value	.06	.94	-.30	.96	-.23	1.01	.55	.81	1.16	.75
Cost value (absence of)	.01	.99	-.02	1.00	-.23	1.03	.14	.90	.22	1.17

Item/factor (per year *z*-scores)	Cluster A		Cluster B		Cluster C		Cluster D		Cluster E	
	M	SD	M	SD	M	SD	M	SD	M	SD
Orientation: mastery	.17	.87	-.32	1.06	-.45	1.13	.54	.67	.82	.63
Orientation: performance	.12	.95	-.30	1.03	-.20	.95	.49	.81	.69	.84
Perceived control	.22	.74	-.26	1.12	-.67	1.25	.36	.80	.67	.85
Perceived control (exams)	.15	.96	-.26	.99	-.53	.98	.40	.79	.87	.84
Study strategy: self-regulation	.06	.93	-.27	.97	-.35	1.06	.59	.77	.97	1.00
Study strategy: control	.08	.92	-.24	1.04	-.30	1.12	.46	.73	.84	.87
Study strategy: memorisation	.10	.89	-.24	1.07	-.31	1.07	.42	.89	.65	.85
Study strategy: elaboration	.04	.91	-.20	1.04	-.21	1.17	.38	.84	.78	.88
Anxiety (absence of)	.12	.91	-.35	.99	-.35	1.04	.58	.76	1.13	.44
Mastery norms (good grade)	-.01	.94	-.11	1.05	-.11	1.14	.38	.77	.56	.88
Subject-comparisons	.09	.96	-.31	1.00	-.28	.97	.56	.78	.98	.58
Peer-comparisons	.07	.96	-.27	.98	-.26	1.09	.50	.83	.89	.68
Social persuasions (praise)	.09	.92	-.32	1.02	-.23	.94	.54	.75	1.05	.81
Vicarious experiences	.04	.95	-.17	1.03	-.28	1.09	.42	.81	.65	.91
Norms/influence (friends)	.03	.92	-.09	1.06	-.09	1.20	.23	.82	.14	1.17
Norms/influence (parents)	.10	.96	-.27	1.01	-.22	1.06	.52	.80	.65	.70
Teacher perceptions	.13	.86	-.24	1.06	-.38	1.20	.31	.87	.81	.77
Teacher/school careers/events	-.01	.96	-.18	1.03	.05	1.10	.42	.83	.50	.98
Cluster size (N, %)	640	42.6%	544	36.2%	114	7.6%	135	9.0%	69	4.6%

Notes: Missing responses were estimated by expectation-maximisation. Means (M) and standard deviations (SD) are reported. Science intentions and their key predictors (identified in ***Section 7****) are shaded in grey.*

Table A9.6: England 2014/2015 (Years 9, 10, 11): task-level accuracy/bias clusters (five clusters), cluster differences

	Overall group difference		Paired group differences, sig. (p)									
Item/factor	Sig. (p)	Size (η^2)	A-B	A-C	A-D	A-E	B-C	B-D	B-E	C-D	C-E	D-E
Science intentions	**<.001**	.126	**<.001**	**.035**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	.063
Task-score	**<.001**	.841	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**.001**
Task-confidence	**<.001**	.446	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Task-conf. accuracy/bias (diff.-score)	**<.001**	.298	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**1.000**	**<.001**	**<.001**	**<.001**
Task sensitivity	**<.001**	.233	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	.624
Task specificity	**<.001**	.126	**<.001**	.178	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Task simple-matching	**<.001**	.166	**.001**	1.000	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	1.000
Task-conf. accuracy/bias (regression-r.)	**<.001**	.152	.055	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**
Self-concept accuracy/bias (regression-r.)	**<.001**	.056	**.035**	**.045**	.539	**<.001**	1.000	**.003**	**<.001**	**.002**	**<.001**	**.001**
Gender (1=boy)	**<.001**	.062	**<.001**	1.000	**.002**	**<.001**	1.000	**<.001**	**<.001**	**.001**	**<.001**	.059
Ethnicity (White)	**<.001**	.062	**.001**	.180	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	.362	.115	1.000
Ethnicity (Black)	.665	.002	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Ethnicity (East-Asian)	**<.001**	.022	1.000	1.000	**.006**	**<.001**	1.000	**.001**	**<.001**	.051	**.002**	1.000
Ethnicity (South-Asian/Indian)	**<.001**	.041	**<.001**	.807	**.044**	**.001**	**<.001**	**<.001**	**<.001**	1.000	.465	1.000
Ethnicity (mixed)	.759	.001	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Ethnicity (other)	.005	.010	1.000	1.000	**.002**	1.000	1.000	**.032**	1.000	.177	1.000	.077
Books at home	**<.001**	.124	**<.001**	**<.001**	.221	**.035**	**.002**	**<.001**	**<.001**	**<.001**	**<.001**	1.000
Parent(s) in science (1=yes)	**<.001**	.018	.090	1.000	**.022**	1.000	1.000	**<.001**	.087	.061	1.000	1.000
Mothers' education	**<.001**	.102	**<.001**	.091	**<.001**	**<.001**	.501	**<.001**	**<.001**	**<.001**	**<.001**	1.000
Fathers' education	**<.001**	.118	**<.001**	1.000	**<.001**	**<.001**	.097	**<.001**	**<.001**	**<.001**	**<.001**	1.000
Current grade	**<.001**	.285	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	.502
Self-concept	**<.001**	.196	**<.001**	**<.001**	**<.001**	**<.001**	.839	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-efficacy	**<.001**	.264	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	.253
Interest value	**<.001**	.174	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**.002**
Utility value	**<.001**	.110	**<.001**	**.003**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	.221
Personal value	**<.001**	.128	**<.001**	**.028**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**

	Overall group difference		Paired group differences, sig. (p)									
Item/factor	Sig. (p)	Size (η^2)	A-B	A-C	A-D	A-E	B-C	B-D	B-E	C-D	C-E	D-E
Cost value (absence of)	**.019**	.008	1.000	.180	1.000	1.000	.437	1.000	.653	**.044**	**.037**	1.000
Orientation: mastery	**<.001**	.121	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	.463
Orientation: performance	**<.001**	.085	**<.001**	**.009**	**.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	1.000
Perceived control	**<.001**	.111	**<.001**	**<.001**	1.000	**.002**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	.264
Perceived control (exams)	**<.001**	.105	**<.001**	**<.001**	.054	**<.001**	.066	**<.001**	**<.001**	**<.001**	**<.001**	**.009**
Study strategy: self-regulation	**<.001**	.111	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	.070
Study strategy: control	**<.001**	.080	**<.001**	**.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	.083
Study strategy: memorisation	**<.001**	.066	**<.001**	**<.001**	**.004**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	1.000
Study strategy: elaboration	**<.001**	.058	**<.001**	.148	**.002**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	.052
Anxiety (absence of)	**<.001**	.149	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**.001**
Mastery norms (good grade)	**<.001**	.033	.860	1.000	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**.001**	**<.001**	1.000
Subject-comparisons	**<.001**	.115	**<.001**	**.001**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**.025**
Peer-comparisons	**<.001**	.092	**<.001**	**.007**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	.064
Social persuasions (praise)	**<.001**	.120	**<.001**	**.008**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**.002**
Vicarious experiences	**<.001**	.051	**.003**	**.014**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	1.000
Norms/influence (friends)	**.006**	.010	.369	1.000	.357	1.000	1.000	**.009**	.645	.120	1.000	1.000
Norms/influence (parents)	**<.001**	.078	**<.001**	**.010**	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	1.000
Teacher perceptions	**<.001**	.077	**<.001**	**<.001**	.461	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**.005**
Teacher/school careers/events	**<.001**	.039	**.029**	1.000	**<.001**	**.001**	.252	**<.001**	**<.001**	**.032**	**.030**	1.000

*Notes: Missing responses were estimated by expectation-maximisation. Significant p-values ($p < .05$) are highlighted in bold. Science intentions and their key predictors (identified in **Section 7**) are shaded in grey.*

Table A9.7: England 2014/2015 (Years 9, 10, 11): intention/attitude clusters (five clusters)

	Cluster A		Cluster B		Cluster C		Cluster D		Cluster E	
Item/factor (per year *z*-scores)	M	SD	M	SD	M	SD	M	SD	M	SD
Science intentions	-.65	.67	.24	.72	1.20	.46	-1.19	.71	1.63	.18
Task-score	-.27	.84	.13	.89	.48	.86	-.39	.85	.72	.79
Task-confidence	-.32	.81	.11	.86	.61	.90	-.79	.87	1.20	.87
Task-conf. accuracy/bias (diff.-score)	.00	.89	-.05	.88	.01	.85	-.29	.93	.29	1.04
Task sensitivity	-.20	.98	.07	.93	.46	.84	-.58	.99	.77	.59
Task specificity	.23	.87	-.10	1.00	-.42	1.02	.60	.75	-.85	.92
Task simple-matching	-.12	.89	-.02	.87	.22	.90	-.08	.81	.46	.89
Task-conf. accuracy/bias (regression-r.)	-.21	.85	.05	.86	.40	.87	-.68	.93	.97	1.03
Self-concept accuracy/bias (regression-r.)	-.30	.85	.17	.83	.57	.80	-.76	1.00	.73	1.17
Gender (1=boy)	-.20	1.01	.08	.98	.07	.99	-.38	1.01	.47	.82
Ethnicity (White)	.25	.85	-.03	.98	-.15	1.03	.34	.81	-.46	1.00
Ethnicity (Black)	-.09	.67	-.04	.89	-.15	.33	-.03	1.05	-.20	.05
Ethnicity (East-Asian)	-.04	.83	-.02	.92	.03	1.14	-.04	.84	.31	1.80
Ethnicity (South-Asian/Indian)	-.22	.76	.09	1.04	.22	1.10	-.32	.67	.45	1.10
Ethnicity (mixed)	.00	1.00	-.04	.90	.01	1.03	.01	1.02	.01	1.02
Ethnicity (other)	-.06	.65	.00	1.01	-.06	.57	-.11	.02	.17	1.59
Books at home	-.17	.99	.14	.94	.27	.98	-.20	1.03	.47	.87
Parent(s) in science (1=yes)	-.07	.94	.02	1.01	.13	1.09	-.25	.78	.31	1.18
Mothers' education	-.20	.94	.09	.99	.22	1.05	-.35	.90	.43	.99
Fathers' education	-.24	.90	.02	.98	.31	1.03	-.46	.90	.60	.98
Current grade	-.24	.91	.19	.93	.52	.88	-.65	.99	1.01	.73
Self-concept	-.36	.79	.25	.76	.79	.76	-.95	.88	1.11	1.14
Self-efficacy	-.30	.91	.25	.85	.61	.71	-.77	1.05	1.09	.40
Interest value	-.43	.64	.39	.58	1.01	.38	-1.59	.83	1.54	.16
Utility value	-.52	.53	.41	.43	1.17	.24	-1.73	.61	1.46	.22
Personal value	-.53	.62	.31	.64	1.07	.47	-1.55	.38	1.73	.18
Cost value (absence of)	.02	.89	-.06	.98	-.13	1.14	.51	1.21	-.19	1.43

	Cluster A		Cluster B		Cluster C		Cluster D		Cluster E	
Item/factor (per year *z*-scores)	M	SD	M	SD	M	SD	M	SD	M	SD
Orientation: mastery	-.21	.90	.13	.87	.78	.54	-.92	1.30	.89	.76
Orientation: performance	-.22	.93	.16	.88	.57	.85	-.70	1.15	.90	.80
Perceived control	-.18	.92	.21	.84	.70	.64	-.86	1.37	1.01	.43
Perceived control (exams)	-.15	.90	.19	.92	.63	.96	-.46	1.14	.60	1.39
Study strategy: self-regulation	-.28	.79	.15	.82	.81	.87	-1.18	1.01	1.18	.93
Study strategy: control	-.26	.91	.16	.78	.80	.72	-.88	1.30	1.10	.93
Study strategy: memorisation	-.26	.92	.15	.87	.68	.83	-1.05	1.13	.77	1.10
Study strategy: elaboration	-.31	.84	.16	.91	.64	.88	-1.11	1.04	1.10	.99
Anxiety (absence of)	-.28	.92	.18	.86	.44	.97	-.61	1.09	.82	1.05
Mastery norms (good grade)	-.09	.97	-.02	1.04	.08	.92	-.13	1.14	.43	.95
Subject-comparisons	-.29	.93	.24	.88	.51	.88	-.68	1.09	.70	1.03
Peer-comparisons	-.23	.91	.14	.89	.41	.94	-.52	1.17	.68	1.11
Social persuasions (praise)	-.37	.85	.25	.79	.71	.83	-.85	.99	1.23	.91
Vicarious experiences	-.11	.93	.18	.88	.50	.82	-.80	1.21	.62	1.17
Norms/influence (friends)	-.16	.91	.12	.96	.40	.91	-.35	1.20	.51	1.23
Norms/influence (parents)	-.46	.77	.36	.69	.94	.49	-1.16	.99	.90	.82
Teacher perceptions	-.29	.89	.20	.83	.65	.76	-.99	1.19	1.32	.47
Teacher/school careers/events	-.28	.89	.14	.96	.38	1.03	-.85	.96	.90	1.16
Cluster size (N, %)	399	34.4%	390	33.6%	171	14.8%	132	11.4%	67	5.8%

*Notes: Missing responses were estimated by expectation-maximisation. Means (M) and standard deviations (SD) are reported. Science intentions and their key predictors (identified in **Section 7**) are shaded in grey.*

Table A9.8: England 2014/2015 (Years 9, 10, 11): intention/attitude clusters (five clusters), cluster differences

	Overall group difference		Paired group differences, sig. (*p*)									
Item/factor	Sig. (*p*)	Size (η^2)	A-B	A-C	A-D	A-E	B-C	B-D	B-E	C-D	C-E	D-E
Science intentions	**<.001**	.623	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Task-score	**<.001**	.131	**<.001**	**<.001**	1.000	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	.477	**<.001**
Task-confidence	**<.001**	.256	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Task-conf. accuracy/bias (diff.-score)	**<.001**	.018	1.000	1.000	**.012**	.152	1.000	.072	**.047**	**.036**	.330	**<.001**
Task sensitivity	**<.001**	.124	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	.179	**<.001**
Task specificity	**<.001**	.132	**<.001**	**<.001**	**.001**	**<.001**	**.002**	**<.001**	**<.001**	**<.001**	**.016**	**<.001**
Task simple-matching	**<.001**	.033	1.000	**<.001**	1.000	**<.001**	**.028**	1.000	**<.001**	**.035**	.510	**<.001**
Task-conf. accuracy/bias (regression-r.)	**<.001**	.161	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Self-concept accuracy/bias (regression-r.)	**<.001**	.196	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Gender (1=boy)	**<.001**	.044	**<.001**	**.026**	.743	**<.001**	1.000	**<.001**	**.033**	**.001**	.051	**<.001**
Ethnicity (White)	**<.001**	.050	**<.001**	**<.001**	1.000	**<.001**	1.000	**.001**	**.006**	**<.001**	.195	**<.001**
Ethnicity (Black)	.279	.004	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Ethnicity (East-Asian)	.114	.007	1.000	1.000	1.000	.088	1.000	1.000	.145	1.000	.604	.210
Ethnicity (South-Asian/Indian)	**<.001**	.053	**<.001**	**<.001**	1.000	**<.001**	1.000	**<.001**	**.036**	**<.001**	.882	**<.001**
Ethnicity (mixed)	.957	.001	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Ethnicity (other)	.199	.005	1.000	1.000	1.000	.393	1.000	1.000	1.000	1.000	.590	.280
Books at home	**<.001**	.045	**<.001**	**<.001**	1.000	**<.001**	1.000	**.003**	.104	**<.001**	1.000	**<.001**
Parent(s) in science (1=yes)	**.001**	.017	1.000	.342	.735	**.037**	1.000	.076	.274	**.013**	1.000	**.002**
Mothers' education	**<.001**	.048	**<.001**	**<.001**	1.000	**<.001**	1.000	**<.001**	.078	**<.001**	1.000	**<.001**
Fathers' education	**<.001**	.078	**.001**	**<.001**	.219	**<.001**	**.010**	**<.001**	**<.001**	**<.001**	.378	**<.001**
Current grade	**<.001**	.176	**<.001**	**<.001**	**<.001**	**<.001**	**.001**	**<.001**	**<.001**	**<.001**	**.002**	**<.001**
Self-concept	**<.001**	.335	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	.065	**<.001**
Self-efficacy	**<.001**	.247	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**.001**	**<.001**
Interest value	**<.001**	.661	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Utility value	**<.001**	.792	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Personal value	**<.001**	.701	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**

	Overall group difference		Paired group differences, sig. (*p*)									
Item/factor	Sig. (*p*)	Size (η^2)	A-B	A-C	A-D	A-E	B-C	B-D	B-E	C-D	C-E	D-E
Cost value (absence of)	**<.001**	.032	1.000	1.000	**<.001**	1.000	1.000	**<.001**	1.000	**<.001**	1.000	**<.001**
Orientation: mastery	**<.001**	.241	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Orientation: performance	**<.001**	.171	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	.127	**<.001**
Perceived control	**<.001**	.227	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	.179	**<.001**
Perceived control (exams)	**<.001**	.109	**<.001**	**<.001**	**.017**	**<.001**	**<.001**	**<.001**	**.015**	**<.001**	1.000	**<.001**
Study strategy: self-regulation	**<.001**	.340	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**.030**	**<.001**
Study strategy: control	**<.001**	.260	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	.219	**<.001**
Study strategy: memorisation	**<.001**	.231	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Study strategy: elaboration	**<.001**	.279	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**.005**	**<.001**
Anxiety (absence of)	**<.001**	.143	**<.001**	**<.001**	**.003**	**<.001**	**.021**	**<.001**	**<.001**	**<.001**	.052	**<.001**
Mastery norms (good grade)	**.001**	.016	1.000	.521	1.000	**.001**	1.000	1.000	**.008**	.688	.177	**.002**
Subject-comparisons	**<.001**	.163	**<.001**	**<.001**	**<.001**	**<.001**	**.017**	**<.001**	**.002**	**<.001**	1.000	**<.001**
Peer-comparisons	**<.001**	.107	**<.001**	**<.001**	**.023**	**<.001**	**.024**	**<.001**	**<.001**	**<.001**	.476	**<.001**
Social persuasions (praise)	**<.001**	.301	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Vicarious experiences	**<.001**	.139	**<.001**	**<.001**	**<.001**	**<.001**	**.003**	**<.001**	**.006**	**<.001**	1.000	**<.001**
Norms/influence (friends)	**<.001**	.062	**.001**	**<.001**	.489	**<.001**	**.022**	**<.001**	**.030**	**<.001**	1.000	**<.001**
Norms/influence (parents)	**<.001**	.448	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	1.000	**<.001**
Teacher perceptions	**<.001**	.292	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Teacher/school careers/events	**<.001**	.167	**<.001**	**<.001**	**<.001**	**<.001**	.062	**<.001**	**<.001**	**<.001**	**.001**	**<.001**

*Notes: Missing responses were estimated by expectation-maximisation. Significant p-values (p < .05) are highlighted in bold. Science intentions and their key predictors (identified in **Section 7**) are shaded in grey.*