INVESTIGATING THE IMPACT OF CLASSROOM CLIMATE ON UK SCHOOL STUDENTS TAKING PART IN A SCIENCE INQUIRY-BASED LEARNING PROGRAMME – CREST

Word count: 15,588

Julie Moote
j.moote@ucl.ac.uk

Institute of Education

University College London

Acknowledgments
The Author would like to thank both Dr Joanne Williams and Professor John Sproule who provided insight, expertise and support that greatly assisted the completion of this research.

ABSTRACT

Understanding students' capacity to direct their own learning in school and beyond has been a sustained area of interest for practising science teachers, policy-makers, and educational researchers. This paper presents results from a longitudinal quasi-experimental study carried out to examine changes in students' self-reported levels of self-regulatory processes and related motivations in science through taking part in an inquiry-based learning programme, the CREST programme. The study included a total of 178 students (aged 12-13) and was conducted in an individual school setting in Scotland with students from one year-group. Previously published and validated measures of self-regulatory processes and related motivations were included as dependent variables and student classroom allocation was used as the independent variable in order to investigate the presence of any differences between the nine CREST classes in terms of changes in self-reported levels on the measured variables. The results showed that overall, students experienced significant increases in levels of self-reported self-regulated learning, self-determination, self-efficacy, intrinsic motivation and personal relevance, and overall science motivation through participation in the CREST programme. By contrast, a reference control class of students not taking part in the programme showed no significant changes in any measured outcomes. The findings documented in this paper also revealed no significant differences between changes in self-reported levels on the outcome variables measured among the nine classes included in this study that participated in the CREST programme. These results highlight the sensitivity of classroom effect studies regarding the choice of analyses and sample size limitations.

INTRODUCTION

Self-regulatory Processes in Science Education

Self-regulation in an academic context involves a series of behaviours (i.e. planning and setting goals and selecting strategies for a learning activity) that individuals cycle through to manage their academic development (Pintrich & De Groot, 1990; Wolters, 2010; Zimmerman, 2002).

Mounting evidence points to the importance of understanding the development of self-regulatory processes with numerous positive impacts on student learning documented (Beishuizen & Steffens, 2011; Dignath & Büttner, 2008; Kistner et al., 2010). Beyond enhancing learning outcomes (Beishuizen & Steffens, 2011), researchers have found that the ability to self-regulate influences students' goal-setting (Ridley, Schutz, Glanz, Weinstein, 1992; Schunk, 1990), increases focus (Zimmerman, 1990) and helps them assess their learning and the effectiveness of any strategies used (Cleary & Chen, 2009). Empirical studies have shown the incidence of poor self-regulation in students today and its impact on academic achievement in a variety of subjects (Matthews, Ponitz, & Morrison, 2009). Self-regulatory and metacognitive processes are not only vital during school but are life-long skills that learners can benefit from after graduation and for self-education later in life (Abdullah & Lee, 2007; Boekaerts, 1997; Kaplan, 2008; Kistner, Rakoczy, Otto, Dignath-van Ewijk, & Büttner, 2010). Considering that life-long learning is at the forefront of both general and science-specific educational reforms (DfE, 2013; Green, 2003, 2011; Hodson, 2003; Reiss, Millar, & Osborne, 1999; Orrow-Whiting, Edwards & Slade, 2007), it is no surprise that fostering these self-regulated processes remains a primary focus of current research and reform initiatives (Beishuizen & Steffens, 2011; Dignath & Büttner, 2008; Kistner et al., 2010; Zimmerman, 2002).

This growing body of research regarding the benefits of developing students' self-regulatory skills have extended to implications for the field of science education (Adey, 1992; Driver, 1989; Driver & Oldham, 1986; White & Frederiksen, 1998; Velayutham, Aldridge, & Fraser, 2012; Zohar, 2004; Zohar & Dori, 2012), due partly to the accumulating understanding of the domain specificity of these constructs and appreciation of the complex nature of science learning (Reif, 2008; Ryder, 2002). The thinking processes necessary for science learning are very distinctive compared to other subjects studied at school (Dillon, 2008; Hodgson & Pyle, 2010; Reif, 2008). Reif argues that one of the main reasons students struggle with learning science is that they approach their learning in science as they would everyday knowledge. In addition, researchers argue that students experience difficulty in learning science due to the demands placed on them to independently accumulate vast amounts of knowledge (eg. De Corte, Verschaffel, & Masui, 2004). As self-regulated learners in science have the ability to control and reflect on their learning, they are generally more motivated and personally interested in the material being studied, show increased academic performance, and are arguably more likely to provide greater contributions to current scientific knowledge (Velayutham *et al.*, 2012; Zohar & Dori, 2012). Together with the downward trends documented in the literature regarding student interest and motivation in science, most threatened between the ages of 10 and 14 years, these findings highlight the relevance of studying these processes in the science subject domain (Archer *et al.*, 2010; Bennett & Hogarth, 2009; George, 2000; Ryan & Patrick, 2001).

The social nature of self-regulation in the science classroom

The literature discussed above relating to conceptualisations of self-regulated learning, highlight the importance of also appreciating the social nature of student self-regulatory processes in science classrooms today. Though intuitively understood as internal by definition, these self-regulatory processes are not entirely intra-psychic, as individuals do not operate autonomously without also being influenced by their social environments (Bandura, 1991). Classrooms are currently understood as social environments (Urdan & Schoenfelder, 2006) and some researchers go as far as to conceptualise different classrooms as different 'cultures' for students (Pintrich, 2003). The classroom climate, defined by Urdan and Schoenfelder (2006) as the general atmosphere in which the learning takes place, can play an important role in developing self-regulated learning and motivation (Vanasupa, Stolk, & Harding, 2010).

Without looking at real classroom settings, generalisations concerning the practical implications of these research findings are of limited value (Martin & McLellan, 2008). The present study therefore builds on the work of researchers who have been developing our understanding of self-regulation in science classrooms (Adey, 1992; Driver, 1989; Driver & Oldham, 1986; White & Frederiksen, 1998; Zohar, 2004; Zohar & Dori, 2012) through evaluating a science education inquiry-based learning activity administered by classroom teachers. In addition, by studying self-regulatory processes in science education, the research presented in this study contributes to the transferability of current self-regulation research findings to other academic domains. This research, therefore, has direct implications for practice and contributes to the identified knowledge gap relating to self-regulation research in natural and specific classroom settings.

As research in the last two decades regarding academic performance has stressed the importance of motivational and cognitive aspects of student classroom learning (Wolters & Pintrich, 1998), it is essential to study these self-regulatory processes and related motivations in adolescent students who were experiencing increased pressure to perform. The physical, mental, and educational transitions experienced by adolescent students further highlights the importance of building on the self-regulation and motivation literature relevant to this age group (Cleary & Chen, 2009; Wigfield & Eccles, 2002). This paper therefore aims to explore students' development of self-regulatory processes through participation in an open inquiry activity that is currently offered as a supplement to the UK science curriculum. The theoretical framework for understanding these self-regulatory and metacognitive processes in the present study will now be presented before describing the intervention programme being studied.

Theoretical Framework

Within the social cognitive perspective, self-regulated learning (SRL) involves several interdependent phases that learners cycle through to manage their academic development (Pintrich & De Groot, 1990; Wolters, 2010; Zimmerman, 2002). Among these phases is the forethought phase, which involves planning and setting goals and selecting strategies for a learning activity. Through the monitoring phase, a student continuously tracks his or her progress and is aware of current performance in relation to the goals that were set. The activities involved in the control phase refer to implementing and acclimating learning strategies to complete the task. Finally, reviewing and responding to the learning experience makes up the reflection phase. In his framework, Pintrich (2004) details the self-regulatory activities involved in each of the phases in four separate areas: cognitive, motivation and affect, behaviour, and context. The present study adopts this multidimensional framework for understanding SRL, and decisions regarding the measurement tools used in this study were guided by this conceptualisation. Velayutham et al. (2012) additionally highlight the importance of implementing strategies to develop self-efficacy and motivations when aiming to promote SRL in secondary school science, and therefore the influence of the CREativity in Science and Technology (CREST) programme on students' beliefs toward their science learning was also investigated. Closely related to the topic of SRL is self-determination, which involves control, choice, and self-initiation of behaviour (Glynn, Taasoobshirazi, & Brickman, 2009). This motivational aspect associated with SRL has been shown to be important in promoting autonomous learning, which helps students retain an intrinsic sense of learning and fosters the development SRL (de Bilde, Vansteenkiste, & Lens, 2011; Deci, Vallerand, Pelletier, & Ryan, 1991). Although not included in many studies of SRL among students, drawing from our previous work (Author, 2016; Authors, 2013), further insight can be gained through incorporating this construct into the understanding of student self-regulatory processes in the present paper. The science education intervention programme investigated will now be presented and positioned within this framework adopted for understanding these constructs among school students.

Inquiry-Based Learning Through a Self-regulatory Lens

Science curriculum reforms in the United Kingdom have seen the implementation of the British Science Association's CREST award scheme. This inquiry-based intervention programme involves a science project (over the course of a minimum of 10 hours—approximately 10 classroom sessions) for students between 11 and 14 years old and is offered to schools as a supplement to the U.K. science curriculum.[i] Each year, over 30, 000 students complete CREST awards, giving them opportunities to explore real-world science, technology, engineering and maths projects within a classroom setting. Led by students and facilitated by teachers, this programme focuses on promoting autonomy and peer collaboration, and on providing students with opportunities to perform self-reflection and self-evaluation. We now consider elements of the CREST programme within the context of self-regulated learning intervention research and connect aspects of the programme to the theoretical framework described above.

To begin the programme, classroom teachers work with students to explore areas of interest and help students formulate scientific questions that they are personally interested in. The programme introduces students to the investigative nature of science, addressing research concerns regarding the development of this appreciation among young students, by providing students the opportunity to come up with their own project hypotheses and methods. Self-regulated learning within the framework outlined above involves goal-directed actions, thoughts, and feelings. Therefore, providing students with opportunities to work towards goals they have set for themselves as part of the CREST programme may also contribute to increases in self-regulated learning, particularly in the forethought stage (Boekaerts & Niemivirta 2000). Research has also shown that participating in open inquiry learning activities, giving students opportunities to be autonomous in their learning and have psychological freedom, can increase autonomous motivation and ownership for learning in science students (Dillon, 2008; Vansteenkiste et al., 2009). Providing students with opportunities for success and ensuring that students find tasks personally meaningful have also been shown to influence the development of self-efficacy and intrinsic motivation (Pintrich, 2003; Schunk & Miller, 2002). As students are provided with the opportunity to choose projects based on personal interests, we predict that CREST may also have a positive impact on the development of these related motivational constructs.

As we have outlined elsewhere (Author 2016; Author et al., 2013), in line with research suggestions in this area, the CREST programme is similar to the Self-Regulation Empowerment Programme (SREP) developed by Cleary and Zimmerman (2004) to foster self-regulated learning in students. Like these initiatives, the CREST programme encourages students to set personal goals, monitor and reflect on their performance processes and outcomes, and make adjustments in order to manage their projects (Cleary & Zimmerman, 2004). Like the above interventions, the CREST programme also directly aligns with De Corte, Verschaffel, and Masui's (2004) Competence, Learning, Intervention, Assessment (CLIA) framework for designing classroom environments that foster self-regulated processes. De Corte and colleagues (2004) identify cooperation among students, active knowledge construction, and self-direction as guiding principles for creating these environments. Therefore, while the programme is not explicitly aimed at developing these self-regulated processes among students, the similarities between CREST and targeted self-regulated interventions, as well as to the CLIA framework. These parallels provide support for this scheme as a viable pedagogical route through which to gain a better understanding of self-regulated processes and related motivations in young science students.

The Present Study- Research Questions

The present study extends our previous work investigating the impacts of the CREST programme on self-regulatory processes and related motivations in young science students (Author 2016; Author et al., 2013). Our 2013 findings supported the predictions that participation in the programme fosters the development of self-reported levels of self-regulated learning and career motivation in science among students at immediate post-test and that these developments are retained at delayed post-test in the months following programme participation. While no significant developments were observed in relation to the other regulatory (self-regulation, cognitive strategies use, and self-determination) and motivational (self-efficacy and intrinsic motivation) constructs studied, contributing to trends in the literature available to date, Author et al. (2013) documented significant decreasing trends in the control group, suggesting the impact of the CREST programme on preventing these potential decreases. Taken together, the results from our previous work highlight the need to replicate the findings in a different sample of students and also to closely examine the impact of participation in the CREST programme at both the class and individual student levels.

Research has shown that teacher and student perceptions of tasks, teacher supportiveness, and social interactions among students are important factors in developing and fostering self-regulation (Patrick, Ryan, & Kaplan, 2007; Ryan, 2000; Vanasupa *et al.*, 2010). As Urdan and Schoenfelder (2006) highlight, the characteristics of the classroom can influence the motivation and cognitions of students in important ways. Therefore, incorporating classroom differences into the research aims of our wider work seems appropriate, and necessary, in order to gain a more complete understanding of these regulatory processes and the student learning taking place through the CREST programme. Further, we appreciate that classroom interventions are (and should be) influenced by the individual teachers implementing them, who come to the classroom with their own values and beliefs (Randi & Corno, 1997). In addition, as our work focuses on understanding these motivations and cognitions through investigating the impact of the CREST programme, focusing additional efforts on the possible influence of different classroom 'cultures' provides an important contribution to knowledge in the field of science education and educational psychology research.

Informed by the research discussed thus far, and the findings of (Author 2016; Author et al. 2013), the study presented in this paper was carried out to gain an understanding of how individual classes of students respond to the CREST programme, focusing on potential class variations. Specifically, the study aims to address four research questions:.

RQ 1: Do classes participating in the CREST programme exhibit different changes in self-reported levels of self-regulated processes and related motivations immediately following participation in the programme compared to the reference control class not taking part in the programme during the course of the study?

RQ 2: Are any developments in the CREST students' self-reports retained four months after participation in the CREST programme for a subset of students?

RQ 3: Are there any differences *between* CREST classes in terms of changes in self-reported levels of the self-regulated processes and related motivations?

RQ 4: If classes are different, does classroom structure at pre-test pertaining to the variables being measured, influence changes in self-reported levels of self-regulated processes and motivation among students participating in the CREST programme?[1]

[1] Research Question 4 will only be addressed if any class differences are found while addressing Research Question 3.

METHODS

Study Design

The present study followed a quasi-experimental design and involved one ‘control’ class and nine ‘CREST’ classes of students from an individual school in Scotland. Students were previously divided into classes based on registration, not ability groups, by the school. Therefore, as students within the year-group were placed into classes and assigned to teachers randomly, the study design is essentially random, which reduces interpretation problems seen in other teacher effect designs (Nye, Konstantopoulos, & Hedges, 2004). All students in the year-group at the school participated in the CREST programme during the academic year except for the one control class. However, due to the small number of the students in the control class (n=18) compared to the total number CREST classes (n=160), this control class will be used more as a reference group. The questionnaire was administered to all students prior to CREST participation, after CREST completion, and again four months after completion to a subsample of CREST students (just over half, n=90) from the nine classes taking part in the CREST programme. Delayed post-tests were not obtainable for the reference control class as they participated in investigations similar to CREST following the administration of the post-test. While a six-month delayed post-test was desired here to replicate the findings of Author et al. (2013), only a four-month delayed post-test was available due to the examination schedules at the school. It should be noted that, contrasting to Author et al. (2013), this four-month delayed post-test was administered following the summer break for this sample of students due to the timing of the CREST programme. This will be considered when interpreting the results presented in this paper.

Participants and Educational Context

Parental consent and child assent were received and data were coded following the ethical guidelines set by [Institution name] and the British Psychological Society. Questionnaires were piloted with 20 students matching the target population (in their second year of secondary school; 12-13 years of age). Items were reviewed by two science teachers and approved for appropriateness and relevance with minor revisions made, and piloted a second time with another 40 students. In order to understand the teaching context, classroom observations were also included in the piloting process, similar to Author et al. (2013). These observations included recording the amount of time spent on the CREST projects, documenting the nature of teacher versus student control, and observing the types of projects conducted. The British Science Association was contacted and information regarding what quality control measures were in place for the CREST programme across schools throughout the UK was obtained. After discussions, it was felt that the teachers involved in the present study administered the programme in very similar ways, supporting the analysis of pre- to post-test change on the key measures. Therefore, no further related teacher data was collected.

Power calculations were performed prior to conducting the present study using the GPower 3.1 programme developed by Faul *et al.* (2009). To observe a medium effect size at an alpha value of .05 and achieve a power of .80, a minimum sample of 190 was required to detect differences between the classes included in this study. Data were therefore collected from 240 students in 12 classes each made up of 20 students. However, only students who completed both the pre-test and post-test were included in the analyses, resulting in an achieved sample of 190 students. According to Green and D'Oliveira (1999), performing inferential statistics on groups of less then 12 participants is inappropriate in psychological research and therefore, two classes were excluded from the analyses due to low numbers of completed pre- and post-tests, leaving 178 students. Similar to Author et al. (2013), this sample size is slightly lower than desired to achieve appropriate power. This will be considered when interpreting the results in the final section of this paper. In addition, as the delayed post-tests were administered to only a subsample of students (n=90), the achieved power in this section of the analyses might have an impact on the results and will also be considered.

As seen in Table 1 below, the gender split between classes in this study was not equal, with males slightly overrepresented. This gender make-up is different to that of Author et al. (2013) and generalisations presented in the discussion at the end of this paper will therefore be made cautiously. In addition, attention should also be drawn to the gender make-up of the reference control class with 13 male students and only 5 female students. This will be considered when discussing the results in the final section of the present paper.

The CREST programme was implemented over the course of eight weeks and students worked on their CREST projects three times a week; completing a total of 24 CREST sessions, each 55 minutes long (total hours on CREST≅22 hours).

-------Insert Table 1 about here-------

Pre- and Post-test Measures

Similar to Author et al. (2013), three self-report measures in the field of self-regulation and motivation were chosen for the present study. These included: the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich & De Groot, 1990); the modified Five Component Scale of Self-Regulation (FCSSR, MacLellan & Soden, 2006); and the Science Motivation Questionnaire (SMQ, Glynn *et al.*, 2009).

As the reliability and implementation of the above measures proved useful in addressing the research questions for our previous work, it was decided that these measures were also appropriate in order to address the research aims of the present study. Using similar measurement instruments will also allow for a cross-study synthesis interpretation, which will be presented in the final section of this paper. Confirmatory factor analyses was performed and the following three tables (Tables 2, 3, & 4) present summaries of the subscales with example items making up each scale. Internal consistency was confirmed by calculating Cronbach's alpha values for each scale based on the sample. The calculated scale reliability results are presented in the following tables alongside the reliabilities reported in the literature for each measure, allowing for comparison.

-------Insert Tables 2, 3, 4 about here-------

Procedure

The questionnaires were administered to all students within one week prior to the CREST intervention and within one week after its completion. Delayed post-tests were administered four months following CREST participation to a subset of participating students (n=90, roughly half of the CREST students included in the study, selected based on timetabling and availability). All questionnaires were administered during class time and students were given 40 minutes to complete them. Completion times ranged between 25 and 35 minutes. Some students did not complete the entire questionnaire and several questions were left blank. Missing data was treated as *user-missing* values and coded appropriately in the statistical software package used (SPSS 19.0). As there were no variables with more than 5% missing data, none were removed from analysis. Missing data analysis was performed similar to Author et al. (2013), and the results of Little's MCAR test was not significant, indicating that data was missing completely at random.

Analysis

One of the strengths of the questionnaire measures used in this study is that they can be treated parametrically and have been extensively used in this manner in the literature (Glynn *et al.*, 2009; MacLellan & Soden, 2006; Wolters & Pintrich, 1998). Therefore, following suit with researchers who are using these measurement tools, and appreciating the increased power and robustness of using parametric analyses (Zimmerman & Zumbo, 1993), the data presented here were also subjected to parametric tests, following the appropriate assumption tests (i.e. homogeneity, normality, and linearity).

Descriptive evaluation involved calculating the means and standard deviations for all subscales in order to determine the variability of scores among the students involved in the study. Preliminary analyses included testing for violations of assumptions of normality and exploring the descriptive statistics to provide further support for parametric treatment of the data. Similar to the work conducted by Chularut & DeBacker (2004), one-way analyses of variances (ANOVA) were also performed in preliminary analyses comparing the ten classes included in this study design on all pre-test measures in order to verify that the classes were matched on pre-test scores.

To address the first research question relating to comparisons of the reference control class to the nine CREST classes, paired-samples *t-tests* were conducted to investigate what changes took place in each of the classes from pre-test to post-test. Literature dating back to the 1960's reported the problem of not directly comparing change scores and assuming *group* differences based only on the above *t-tests*. Researchers state that when intact classrooms are assigned to treatments, the paired-samples *t-tests* have too small an error term due to the fact that randomisation is lumpy (Campbell & Stanley, 1966; Zimmerman, 1997). Therefore, following the suggestions of Campbell & Stanley (1966), and more recently Zimmerman (1997), further independent samples *t-tests* were also conducted to directly compare the change scores of each of the nine CREST classes to the reference control class in order to explore the presence of *group* differences. In order to investigate whether any developments in students' levels of self-regulated processes and the related motivational factors studied were retained four months after participation in the programme, student scores on the variables at post-test and delayed post-test were compared through conducting a series of similar paired-samples *t-tests*. As delayed post-test data was only available for a subset of the CREST participants (n=90), only these students were included in this section of analysis.

While both multilevel and regression analyses were considered for this study in order to address the third research question relating to class differences, due to the total sample size as well as the small sample size within each class, the analytical strategy chosen for this study was to look at the differences between classrooms in experience within the CREST programme. Classroom effects in this study are therefore operationalised as between-classroom differences in change scores from pre-test to post-test. As the analysis of change scores addresses group differences, this was the method executed in this study (Smolkowski, 2013; see Author et al., 2013 for a more detailed justification).

The analyses reported in this study involve conducting a series of multiple tests, which raises concern of Type 1 error. While conservative Bonferroni adjusted alpha levels have been adopted to control for this, we additionally focus on the interpretation of effect sizes rather than *p* values to reduce the risk of Type 1 errors and provide insight into the practical implications of the magnitude of the reported differences (Cumming et al. 2012; Wasserstein & Lazar, 2016). Cohen's d statistic will accompany any t-tests presented, partial eta squared ($\eta\rho2$) values will be included to represent effect sizes for

ANOVAs (see Table 5 below) and Cramer's V for chi-squared tests. For details on effect size calculations and interpretations see Pallant (2010). Effect sizes are also reported in addition to significance values, as reporting and interpreting effect sizes can provide insight into the practical implications of the magnitude of reported differences (Field, 2013). As withour previous work, Cohen's d statistic will accompany any *t-tests* presented and partial eta squared ($\eta\rho^2$) values will be included to represent effect sizes for ANOVAs.

-------Insert Table 5 about here-------

In order to simplify the analyses and make results more manageable, the results will be presented in sections relating to the research questions outlined earlier. Table 6 below provides an overview of these results sections with an outline of the corresponding parametric analyses that were conducted.

-------Insert Table 6 about here-------

RESULTS

Preliminary Analysis: Exploring Class Pre-test Differences on All Measured Variables

One-way ANOVAs were conducted to explore differences between classes on all pre-test variables measured. This step of the analysis was necessary in order to ensure the validity of any further tests of intervention effects. Considering that sample sizes across groups were slightly different, Gabriel's procedure was used following the suggestions of Field and Hole (2003). The Games-Howell procedure was also run following the recommendations of Field (2009, 2013), due to the uncertainty of knowing whether population variances are equivalent. The results from the one-way ANOVAs showed no significant pre-test differences between the 10 classes included in this study on any pre-test variables using a conservative Bonferroni adjusted alpha level (p=.01) (to reduce the risk of Type 1 error). From these preliminary results, it is reasonable to suggest that any systematic differences in the variables being studied at post-test and delayed post-test are likely due to either class or teacher effects through administering the CREST programme. Preliminary analyses also involved exploring gender differences throughout the data. As no significant differences were found between boys and girls on all pre- and post-test measures, the data were combined for the two genders in all subsequent analyses conducted in the present study.

RQ 1: Investigating Group Differences in Pre- to Post-test Change

Before describing the experience of the nine CREST classes, the reference control class results will be presented. To explore changes from pre-test to post-test in students' self-reported levels of self-regulated processes and related motivations in the reference control class that did not take part in the CREST programme, paired-samples *t-tests* were conducted on all variables measured. Results showed that the control class experienced no significant changes from pre-test to post-test on any measured outcome variables. These non-significant results, as shown in Table 7, suggest that the students not taking part in the CREST programme reported no significant increases in the self-regulated processes and related motivations measured in this study.

-------Insert Table 7 about here-------

The paired-samples *t-tests* presented above were repeated for each of the nine CREST classes included in the present study separately, and taken together in order to investigate the significance of any measured changes from pre-test to post-test. While it may seem counter-intuitive to consider all CREST classes together when looking at pre- to post-test change in a study investigating *class* differences, as the research question here relates to *group* differences, looking at the average change from pre-test to post-test for all CREST classes together may provide helpful insight. In addition, looking at the overall change considering all students participating in the CREST programme together, helps connect to the results presented in Author et al. (2013) relating to *group* differences, replicating the results with a much larger sample of students. However, it is important to clarify at this point that the control class is only being considered as a reference class and no direct statistical tests are being conducted, as comparing 160 CREST students to the 18 students not taking part in the programme would be inappropriate.

While an alpha level set to .05 may not be expected due to the number of similar tests run in this section of analysis, the results will be carefully interpreted appreciating their practical significance. As the sample size for each class included in the present study design is very small, the likelihood of finding significance is already very low (Feise, 2002; Koretz, 2005). Therefore, if the alpha value is adjusted to reduce the chances of Type 1 error, the chances of Type 2 errors occurring may increase. All results presented below had a medium to large effect size (reported in the Tables below), unless otherwise mentioned and therefore will be considered as significant intervention effects.

Tables 8, 9, 10, & 11 show the means and standard deviations for each of the nine CREST classes at pre-test and post-test. The computed change scores as well as the significant paired-samples *t-test* results are also presented in the following tables. As no significant changes were found in the reference control class (see Table 7 presented previously), the presence of significant changes from pre-test to post-test in any of the individual nine CREST classes may indicate an intervention effect for the CREST programme. The following four tables are organised into self-regulated processes, related motivations, test anxiety, and science-specific motivational outcome measures.

Self-Regulated Processes

From Table 8, it can be seen that Class 7 experienced significant increases in self-regulation measured on the MSLQ from pre-test (M=4.20, SD=.856) to post-test (M=4.56, SD=.936) with the Cohen's d value indicating a medium effect size. Class 2 also showed significant increases from pre-test (M= 2.66, SD= .330) to post-test (M=3.04, SD=.553) in self-regulated learning measured by the FCSSR with the Cohen's d statistic (d=.84) indicating a large effect size. Significant increases were also found in Class 3 on the SMQ measure of self-determination. Results from the paired-samples *t-test* showed that Class 3 experienced significant increases from pre-test (M=3.57, SD=.727) to post-test (M=3.92, SD=.647) with the Cohen's d value indicating a medium effect (d=.50).

These significant findings demonstrate that, unlike the control class, some of the CREST classes experienced significant changes in self-regulation, self-regulated learning, and self-determination following participation in the programme suggesting the possibility of intervention effects for the CREST programme.

Further support was found for the above results relating to self-regulated processes when considering the nine CREST classes together. The results of the paired-samples *t-tests* conducted for the CREST classes taken together were significant for SMQ self-determination, with the CREST classes on average increasing from pre-test (M=3.62, SD=.797) to post-test (M=3.76, SD=.747). The mean increase in self-reported levels of self-determination was .134 with a 95% confidence interval ranging from .0388 to .229 and the Cohen's d effect size statistic (.18) indicating a small, approaching medium effect. While average changes on the self-regulated learning measure did not show significance at the conservative alpha level of .01 (to control for Type 1 error), a closer look into the effect size of this test highlights the potential practical utility of the result. Table 8 shows that the effect size for the increase in self-reported levels of self-regulated learning in the nine CREST classes considered together as a *group*, from pre-test (M=2.44, SD=.615) to post-test (M=2.53, SD=.642), according to Cohen's d statistic (.14), was small.

Together, these results provide support for the overall impact of the CREST programme on the students taking part. They also contribute further evidence for the possibility that, while only three classes appeared to demonstrate significant changes on some self-regulated process measures following CREST participation, the overall trend in the nine classes taking part in the programme on these outcome measures was an increasing one. Overall, the results presented in this section provide support for the the impact of CREST programme participation on self-regulated processes and also externally validate the results presented previously in Author et al. (2013) by replicating them in a difference student sample.

-------Insert Table 8 about here-------

Related Motivations

Table 9 presents a summary of the pre- and post-test means and change scores for the nine CREST classes, as well as any significant paired-samples *t-test* results for the related motivational variables. From Table 9, it can be seen that significant increases were observed in Class 2 relating to both measures of self-efficacy. On the MSLQ measure of self-efficacy, Class 2 showed significant increases from pre-test (M=4.88, SD=.927) to post-test (M=5.44, SD=1.10) with the Cohen's d value for this test (d=.55) indicating a medium effect size. Similar significant increases from pre-test (M=3.64, SD=.683) to post-test (M=3.97, SD=.816) were also noted on the SMQ measure of self-efficacy with the Cohen's d value also indicating a medium effect size (d=.43). While no other CREST classes demonstrated significant increases in self-efficacy, increasing trends were observed with no significant decreases noted. These results suggest that the CREST classes experienced significant changes in self-efficacy while the reference control class did not (as shown previously in Table 7).

Providing support for the observed trends described above, Table 9 also shows that, overall, the CREST classes experienced significant increases in self-reported levels of self-efficacy from pre-test (M=3.50, SD=.854) to post-test (M=3.70, SD=.788) on the SMQ at the adjusted alpha level of .01. The mean change score was .202 with a 95% confidence interval ranging from .105 to .299 and the Cohen's d statistic (.23) indicated a small, approaching medium effect. A similar increasing trend (not significant at the adjusted alpha level) was found for self-efficacy measured on the MSLQ when all classes were considered together and the Cohen's d statistic for this test (d=.17) indicated a small, approaching medium effect.

Relating to intrinsic motivation, Table 9 shows that Class 2 experienced significant increases on the SMQ measure of intrinsic motivation and personal relevance with the post-test score (M=4.08, SD=.620) being significantly higher then the starting pre-test score (M=3.34, SD=.445) measured before participation in the CREST programme. The Cohen's d value for this test documented a large effect size (d=.81). While no other classes demonstrated significant increases in intrinsic motivation, similar to self-efficacy, increasing trends were noted. The observation of these increasing trends for intrinsic motivation and personal relevance measured by the SMQ were supported by the analysis of the CREST classes considered together.

Results from the paired-samples *t-tests* considering all nine CREST classes together showed that the classes, on average, increased significantly from pre-test (M=3.48, SD=.745) to post-test (M=3.63, SD=.718) on levels of self-reported intrinsic motivation and personal relevance measured on the SMQ. The mean change score for this variable was .153 with the 95% confidence interval ranging from .0711 to .235 and the Cohen's d statistic (.21) indicating a small, approaching medium effect. Similar to the self-efficacy results presented above for the MSLQ, paired-samples *t-tests* for intrinsic motivation measured on the MSLQ did not reach significance at the conservative alpha level of .01.

The results presented above relating to self-efficacy and intrinsic motivation provide support for the presence of *group* differences in this study relating to these constructs. As significant changes in self-efficacy and intrinsic motivation were reported in some of the CREST classes, with no significant changes reported in the reference control class (see Table 7 reported earlier), the presence of intervention effects for the CREST programme is possible regarding these related motivations.

The above results provide further support for the first research question in regards to the specific trends expected on these outcome measures following CREST participation. While it was predicted that levels of self-efficacy and intrinsic motivation would not decrease following participation in the CREST programme, the above results suggest that for this sample of students, unlike in Author et al. (2013), self-reported levels of self-efficacy and intrinsic motivation, on average, increased following participation in the programme. However, these results are interpreted with caution as the absence of a control group matching the size of the CREST group for this study (n=160) limits the conclusions that can be drawn from these findings alone.

-------Insert Table 9 about here-------

As we previously reported significant increases in test anxiety for students participating in the CREST programme, similar significant increasing trends were predicted here. However, Table 10 below presents the results from the paired-samples *t-tests* conducted for test anxiety, which showed no significant changes in any of the nine classes taking part in the CREST programme with the overall trend also indicating no significant changes in this outcome measure. In contrast to the findings documented in Author et al. (2013), these results suggest that participation in the CREST programme did not have a significant impact on students' self-reported anxiety levels towards taking tests. This result is further supported by the non-significant results reported earlier in Table 7 for the reference control class. However, an inspection of the means and change scores shown in Table 10 below relating to test anxiety indicates an increasing trend which will be considered when interpreting these results in the discussion at the end of this paper.

-------Insert Table 10 about here-------

Science-Specific Motivations

Table 11 shows the means and standard deviations for the science-specific motivational measures. No significant changes from pre- to post-test were reported in any of the nine CREST classes in terms of self-reports of career motivation. While an increasing trend was observed, the paired-samples *t-test* conducted for career motivation with all CREST classes considered together was not significant and is also shown in Table 11. These results are not in line with the findings of Author et al. (2013) which showed that participating in the CREST programme significantly increased students' self-reports of motivation for pursuing science careers from pre-test to post-test. From the results presented in Table 11, and the non-significant reference control class results presented earlier in Table 7, it can be concluded that no differences between the groups were found relating to this outcome measure.

As outlined earlier, it was predicted that grade motivation in science and overall science motivation would increase alongside any increases in related motivations. For grade motivation in science, significant *decreases* were found from pre-test (M=4.03, SD=.447) to post-test (M=3.77, SD=.601) in Class 7 with the Cohen's d statistic indicating a medium effect size (d=.57). While significance was only found in this class, decreasing trends were observed in most CREST classes. Considering the non-significant results presented earlier in Table 7 for the reference control class, these results suggest the possibility of *group* differences relating to grade motivation.

Supporting the observed decreasing trends, the results from the paired-samples *t-tests* conducted on all nine CREST classes taken together showed that for grade motivation, self-reports at pre-test (M=3.84, SD=.642) were significantly higher than self-reports at post-test (M=3.73, SD=.622). The mean decrease in self-reported levels is shown below in Table 10 and the 95% confidence interval ranged from -.0247 to -.217 with the Cohen's d statistic (.19) indicating a small, approaching medium effect size. As grade motivation was not included in Author et al. (2013), no results were available to be replicated. However, it was expected that in addition to increasing levels of career motivation in science, the CREST programme would also have a positive influence on students' self-reports of their motivations to attain high grades in science class. The results presented above relating to grade motivation therefore suggest that participation in the CREST programme led to decreases in students' motivations for achieving high grades in science. These results will be discussed further at the end of this paper.

Relating to overall science motivation measured by the SMQ, Class 1 showed significant increases from pre-test (M=98.2, SD=15.7) to post-test (M=103.1, SD=15.6) with the Cohen's d statistic indicating a medium effect size (d=.32). Similar results were found in Class 2, with significant increases from pre-test (M=111.2, SD=9.58) to post-test (M=119.4, SD=6.80) and with Cohen's d indicating a large effect size (d=.99). Together with the non-significant changes in the reference control class presented in Table 7 earlier, these results provide support for the presence of *group* differences relating to overall science motivation following CREST participation. When investigating changes for all CREST classes considered together, for total science motivations as measured by the SMQ, pre-test self-reports (M=103.16, SD=17.3) were found to be significantly lower than post-test values (M=106.72, SD=16.65). The mean change score was 3.56 with the 95% confidence interval ranging from 1.85 to 5.26 and the Cohen's d statistic (.21) indicating a small, approaching medium effect. These results suggest that participating in the CREST programme has a positive impact on students' overall motivations for their science learning.

-------Insert Table 11 about here-------

Validation of Intervention Effects Relating to RQ 1

The results presented, thus far, provide some support that the CREST classes experienced different changes on several measured outcomes compared to the reference control class included in this study. However, in order to verify the presence of any intervention effects and *group* differences, independent-samples *t-tests* directly comparing the change scores of each CREST class to the reference control class were performed when significant increases were found from pre-test to post-test in a CREST class. Table 12 shows the significant differences that were found for these CREST class versus reference control class change score comparisons.

Significant differences were found between the MSLQ self-regulation change scores of Class 7 (M=.360, SD=.612) and the reference control class (M=-.556, SD=.566) with the Cohen's d value indicating a large effect size (d=.726). Together with the significant increases seen from pre-test to post-test in Class 7 on MSLQ self-regulation in the previous results section (reported in Table 7 earlier), these results provide further support for the impact of the CREST programme on self-regulation in this class. Significant differences were also found on FCSSR self-regulated learning between CREST Class 2 (M=.381, SD=.467) and the reference control class (M=-.0037, SD=.359) with the Cohen's d value again indicating a large effect size (d=.924). While significant pre- to post-test change was noted in Class 2 on this variable in the previous analysis, the findings presented in Table 12 provide further support for the effectiveness of the CREST programme in increasing self-regulated learning in this class.

The results of the independent-samples *t-test* comparing the MSLQ self-efficacy change scores of Class 2 (M=.556, SD=.916) to the reference control class (M=-.0432, SD=.484) were statistically significant with the Cohen's d value indicating a large effect size (d=.818). Significant differences were also found for the grade motivation self-reports of Class 7 (M=-.316, SD=.658) and the reference control class (M=.0625, SD=.418) with the Cohen's d value indicating a large effect again (d=.691). These results, therefore, provide further support for the *group* differences noted earlier regarding these outcome variables

No significant differences were found in the other CREST versus reference control class comparisons performed for the significant paired-samples *t-tests* reported earlier in Tables 8 through to 11. In other words, while some classes experienced significant changes in self-determination, overall science motivation, and intrinsic motivation from pre- to post-test, when directly compared to the changes experienced in the reference control class, no significant *group* differences were found on these variables. This section of results validating the changes reported above relating to the presence of group differences highlights the importance of directly comparing the change scores among the classes included in this study.

-------Insert Table 12 about here-------

RQ 2: Retention Effects for the CREST Students

In order to investigate the presence of any retention effects for developments in self-reported levels of the self-regulated processes and related motivations measured, paired-samples *t-tests* were conducted on the subset of students that completed the four-month delayed post-tests. This subset included 90 students from the 160 students participating in CREST (56% of the CREST students included in the study).

Table 13 shows the means and standard deviations of the post-tests and delayed post-tests as well as a summary of the paired-samples *t-test* results. The only test presented in Table 13 to reach significance was intrinsic motivation and personal relevance measured by the SMQ. Students that completed the delayed post-tests showed significant decreases in their self-reports of intrinsic motivation and personal relevance from post-test (M=3.50, SD=.70) to delayed post-test (M=3.33, SD=.67) with the Cohen's d value indicating a small approaching medium effect size (d=.249). This result will be discussed in relation to published literature documenting decreasing trends in these intrinsic motivations.

The absence of any other significant changes in the measured variables suggests that the changes observed in the outcome measures reported in the results sections above were retained four months following CREST participation. These results also suggest that no consolidation of effects occurred in the months following programme participation. In other words, on the outcome measures that showed no immediate group differences or intervention effects (cognitive strategies use, test anxiety, and career motivation), no delayed impact of the intervention was later reported.

-------Insert Table 13 about here-------

RQ 3: Investigating Class Differences in Response to the Intervention

The results presented earlier relating to the first research question demonstrated the presence of *group* differences between the classes taking part in CREST and the reference control class. As some of the CREST classes did not experience significant changes from pre-test to post-test while others did, the above findings suggested the possibility of *class* differences regarding the effect of the programme on the students participating. However, as with the direct comparisons made earlier in order to validate the research results relating to the first research question of the present study, similar direct comparisons need to be made in order to address the third research question relating to *class* differences in terms of changes in self-reported levels of the outcome measures following participation in the CREST programme. While it may seem that the results presented in this section under-cut some of the previous findings presented, what this additional analyses provides is clear insight into the influence of the statistical approach adopted on the research findings in studies investigating class differences in response to interventions.

In order to address the third and final research question of the present study, a series of one-way between-groups MANOVAs and ANOVAs were performed to investigate differences among the nine CREST classes regarding changes in self-reported levels of the self-regulated processes, related motivations, and science-specific motivations measured in this study. Table 14 presents a summary of these results.[2] From Table 14, it can be seen that all of the analyses of variance results reported in this section indicate small, approaching medium effect sizes.

For self-regulated processes, similar to the MANOVA analysis presented in Author et al.

[2] All pre-test, post-test, and change score means and standard deviations were presented previously in Tables 8, 9, 10, and 11.

(2013), four dependent variables were used: MSLQ self-regulation, MSLQ cognitive strategies use, FCSSR self-regulated learning, and SMQ self-determination change scores (post-test minus pre-test). As the third research question of this study investigates whether the classes participating in the CREST programme change in levels of the outcome measures to the same extent, the independent variable here is *class*. As shown in Table 14, no statistically significant differences were found between the nine CREST classes on the combined dependent variables for the self-regulated processes. Relating back to the third research question regarding the presence of *class* differences and building on the analyses presented thus far in the paper, these results suggest that the nine classes participating in CREST experienced similar changes in their self-reports of the self-regulated processes measured in this study. Remembering the findings discussed earlier relating to the first research question, these results suggest that while no differences between classes regarding the changes from pre-test to post-test were observed, on average, students taking part in CREST showed significant increases in self-reported levels of self-regulated learning and self-determination. These results were not expected in the present study.The multivariate results investigating class differences in self-efficacy scores on the MSLQ and SMQ (presented in Table 14) showed no statistically significant differences between the classes on the combined dependent variables. Similar non-significant results were also found on the multivariate tests for intrinsic motivation change scores on both the MSLQ and SMQ and are also presented in Table 14. As with self-efficacy and intrinsic motivation, the multivariate results investigating class differences in test anxiety change scores on the MSLQ and SMQ showed no statistically significant differences between the classes on the combined variables (see Table 14). From the results presented above relating to these constructs, it can be interpreted that the nine classes taking part in the CREST programme experienced similar changes in their self-reports of these related motivations answering the third research question. Recalling the results presented earlier in relation to the first research question of the present study, these results also suggest that while the nine CREST classes experienced similar changes in these variables, on average, they showed significant increases in self-efficacy and intrinsic motivation, with no statistically significant changes from pre-test to post-test reported on either of the test anxiety measures.

A series of one-way ANOVAs comparing the change scores of the nine CREST classes was also performed investigating class differences in the science-specific motivations measured on the SMQ; career motivation, grade motivation, and overall science motivation. As shown in Table 14, there were no statistically significant differences between the nine CREST classes regarding these three science-specific motivations measured. This suggests that the nine classes did not differ in terms of the impact of the CREST programme. Again recalling the results presented earlier relating to the first research question, together these results suggest that the nine CREST classes were no different in the pre-test to post-test changes they experienced and, on average, they experienced no changes in career motivation, decreases in grade motivation, and increases in overall science motivation.

-------Insert Table 14 about here-------

DISCUSSION

This study aimed to understand the impact of participating in the CREST programme on changes in self-reported levels of self-regulated processes and related motivations, extending the results presented in Author 2016 and Author et al. (2013). This study was also interested in looking at the experience at the classroom level in order to investigate the presence of differences between the nine CREST classes in regards to the changes observed on the outcome measures included. The results will now be discussed in relation to the three research questions.

RQ 1 & 2: Exploring the Experience of CREST Classes Compared to the Reference Control Class and Retention Effects

Self-Regulated Processes

The present study aimed, firstly, to explore the differences between changes in self-reported levels of self-regulated processes and related motivations in classes taking part in the CREST programme to a reference control class of students not participating in the programme. The results presented provide some support for *group* differences regarding changes in self-reported levels of the key self-regulated process outcome measures. Results showed that while the reference control group did not experience any significant changes in self-reported levels of any variables measured, significant changes from pre-test to post-test were found in some of the CREST classes on self-regulation, self-determination, as well as self-regulated learning. However, when verifying these results by comparing the change scores of the classes that experienced significant changes along the outcome measures to the reference control class, significant differences were only found on the self-report measures of self-regulated learning and self-regulation in two classes. These results support the claims of researchers who state that the paired-samples *t-tests* have too small an error term due to the fact that randomisation is more lumpy in studies utilising natural classroom structure (Campbell & Stanley, 1966; Zimmerman, 1997). The conclusion that can be drawn from these results is that *group* differences were found for self-regulation and self-regulated learning, but only in two of the nine classes taking part in the CREST programme. However, as the classes involved in this comparison include very small sample sizes, conclusions are drawn cautiously.

As the above analyses only found significant group differences on self-regulation and self-regulated learning for two classes, it could be concluded that no other group differences were present. However, as the analyses described earlier included looking at *all* students participating in CREST, the results also indicated that on average, students taking part in CREST showed significant increases in self-reported levels of self-determination and self-regulated learning. These results suggest that while only a few classes showed significant pre-to-post-test change on the self-regulated process variables measured, with only two of these classes being significantly different from the reference control class, general increasing trends were noted among the CREST participants. While no comparisons can be made between these trends and the reference control class (due to large differences in sample sizes), comparing these results to other research findings may provide further support for group differences in self-determination and self-regulated learning.

Considering the downward trends over the school term documented by Berger and Karabenick (2011) and discussed in (Author et al. 2013), the practical significance of the overall increases in self-regulated learning and self-determination reported among students taking part in the CREST programme in this study are highlighted. Therefore, together with the documented literature findings and the results presented in (Author et al., 2013), the results here provide strong support for the practical utility of investigating the CREST programme within a self-regulated learning framework and exploring the impact of participation in the programme on these regulatory processes in students.

While significant changes were found for self-regulated learning and self-determination, no significant changes were observed in self-reported levels of cognitive strategies use for the students participating in the CREST programme. However, relating these findings to published research, as well as to the results presented in Author et al. (2013), provides additional insight into how to interpret these results. Van der Veen and Peetsma (2009) conducted a study with similarly aged students and found that levels of cognitive strategies use, as measured by the MSLQ, decreased over the course of the school year. Through looking at the means presented earlier in Table 8, it can be seen that while some classes taking part in CREST did experience small decreases in levels of self-reported cognitive strategies use throughout the course of the study, the majority of CREST classes experienced increases in self-reported levels of cognitive strategies use. In addition, while some of the classes showed decreasing trends, it is important to note that overall, no significant decreases were reported among the students taking part in the CREST programme. These results suggest that while CREST does not involve explicit cognitive strategies instruction, participation may help students maintain pre-test levels of self-reports of this construct. Through relating these findings to the work of Van der Veen and Peetsma (2009) and appreciating that they replicate the results presented in Author et al. (2013) it seems possible that the CREST programme prevented the decreases in levels of perceived cognitive strategy use that may have occurred without intervention.

Graham and Harris (1993) highlight that self-regulated process strategies need to be maintained over longer periods of time and be transferable to other classroom-learning subjects. Similar to Author et al. (2013), delayed post-tests were also incorporated in the present study design in order to investigate the presence of any long-term impacts of CREST programme participation. Relating to all self-regulated process variables, the results of this study are in line with our previous findings in that no significant changes in self-reported levels were observed from post-test to delayed post-test. From these results, it can be interpreted that any developments documented in these outcome measures (increases in self-regulated learning and self-determination) were maintained at the four-month post-test for the students that took part in the CREST programme.

However, similar to Author and colleagues (2013), as no delayed post-tests were obtained from the reference control group and that the subset available for the delayed post-test was essentially a convenience sample of (n= 90) CREST students (which also limits power), the conclusions here are drawn cautiously.

Related Motivations

Other significant changes in self-reports measured in the present study were found for self-efficacy and intrinsic motivation and personal relevance. These findings showed significant increasing trends in self-efficacy and intrinsic motivation for students taking part in the programme but are, however, different from the findings of Author and colleagues (2013). Our previous work showed no significant changes in levels of self-reported self-efficacy and intrinsic motivationwhich was not in line with the predictions made. We will now move on to examine the significant increases in self-efficacy and intrinsic motivation documented in the present study within the context of similar studies reported in the literature.

In their work, described earlier in relation to the self-regulated process results for the present study, Berger and Karabenick (2011) also investigated changes in self-efficacy and intrinsic motivation over the course of the four-month study which was conducted in the United States. Their results showed that while self-efficacy remained stable over the course of the school term for students in mathematics, self-reported levels of intrinsic motivation decreased. There seems to be a general consensus in the literature that self-efficacy and intrinsic motivation decrease, or remain stable over a school term or a course at college, with little evidence of increasing trends without pedagogical intervention (Chase, 2001; Fredricks & Eccles, 2002; Gao, Lee, Solomon, & Zhang, 2009; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Pintrich & Schunk, 2002; Ryan & Deci, 2000; Zusho, Pintrich, & Coppola, 2003). In line with the research trends discussed here, self-reported levels of self-efficacy and intrinsic motivation remained stable in the reference control class over the course of the study.

In contrast, recent research findings have documented increases in self-efficacy and intrinsic motivation in students following participation in interventions aimed at developing self-regulated learning (Fuchs *et al.*, 2003; Stoeger & Ziegler, 2010). Giving students more responsibility and choice, and providing them with opportunities to plan and evaluate their learning, builds self-confidence and can help maintain high levels of self-efficacy and intrinsic motivation (Patall, Cooper, & Wynn, 2010; Schunk & Ertmer, 2000; Zimmerman, 2000). The results documented here are in line with these literature findings as CREST students, on average, showed significant increases in their self-reported levels of these constructs.

Results presented in this paper also indicated that the developments seen in self-efficacy described above were retained four months following CREST programme participation. However, regarding intrinsic motivation measured on the SMQ, significant decreases were found for all CREST students on average from post-test to delayed post-test. These results suggest that while intrinsic motivation was higher following participation in the programme, these developments were not retained four months later after the summer break. Considering the decreasing trends presented in the literature discussed earlier, it is possible that these significant decreases would also be seen in a control group of students not taking part in the programme with potentially more dramatic decreases observed. However, these results do suggest that measures may need to be in place in order to build on the motivational developments for students taking part in the programme.

In the present study, no significant changes were found in self-reported levels of test anxiety for the reference control class and the nine CREST classes. One possible explanation for the lack of significant changes in test anxiety could be due to the fact that the temporal interval implemented in the present study design (eight weeks) was too brief to detect intervention effects (Berger & Karabenick, 2011). However, as we previously investigated changes over the course of only five weeks and documented significant increasing trends in self-reported levels of test anxiety for students participating in the CREST programme (Author et al., 2013), this explanation seems unlikely.

While no significant *group* differences were found relating to test anxiety at the adjusted alpha level of .01, increasing trends were noted among the CREST classes. As discussed in Author et al. (2013), Rozendaal and colleagues (2005) similarly found that self-regulated learning-based innovation programmes may not be able to solve student problems with anxiety. As Rozendaal and colleagues highlighted that various forms of anxiety are detrimental to the development of self-regulatory skills, further research is needed in order to understand the impact of CREST programme participation on student levels of self-reported test anxiety.

It is also important to discuss the results for the related motivational constructs above in terms of the different measurement instruments used in the present study. As with Author et al. (2013), the present study included two measures for each of the related motivational constructs (self-efficacy, intrinsic motivation, and test anxiety) and different results were obtained in terms of statistical significance. While the SMQ measures showed mostly significant changes, the MSLQ did not. One possible explanation for these results could be that while the MSLQ can be formatted to be subject specific, the SMQ was specifically designed for implementation in science classrooms. These results have important implications in terms of measurement issues and also researchers' choice of which measures to include in their work.

Science-Specific Motivations

The findings presented above relating to the science-specific motivations showed significant increases in overall science motivation and while not significant, an increasing trend for career motivation in science was also found. Taking part in the CREST programme allows students to experience hands-on science and motivates them generally, but also specifically to increase their interest in pursuing science careers, and these developments are retained four months following programme participation.

The results presented in this paper showed significant decreases for the CREST classes, on average, for grade motivation in science. Group differences were also found between the reference control class and CREST Class 7. This may be explained by the timing of the study as students and teachers were focusing on completing the CREST programme with no unit tests coming up. During the CREST programme, as classroom teaching time is devoted to completing the projects, it is possible that this lack of anticipation for an upcoming test could explain the decrease in grade motivations seen. It is also possible that participation in the CREST programme helps students understand that science education is about more than achieving good grades however, further research is needed to explore these possibilities. Considering recent shifts regarding the importance placed on grades by educators and policy makers, as well as the detrimental effects of students being motivated solely by external rewards, highlights the practical implications of these findings relating to grade motivation (Kohn, 2011).

RQ 3: Investigating Class Differences in Response to the CREST Programme

Another aim of the study presented in this paper was to investigate *class* differences regarding changes in self-reported levels on the key variables measured. From the paired-samples *t-test* results presented above for the nine CREST classes, one conclusion that could be drawn is that, as some of the classes experienced significant changes while others did not; the classes were different regarding their responses to the CREST programme. However, in order to directly compare these differences, ANOVA analyses were conducted on the pre-test to post-test change scores on all outcome measures.

The overall non-significant ANOVA results comparing the change scores of the nine CREST classes suggest that these differences were not large enough to reach statistical significance. From the CREST class ANOVA comparison results presented in this study, the conclusion that no class effects were found on any of the variables could be inferred. Another possibility is that while the other classes were not significant, they were reaching significance and overall, classes on average, increased. These results may explain the overall non-significant multivariate results presented in the previous section relating to class differences in change scores. From this perspective it could be concluded that no class differences were present and, on average, students developed self-regulated learning, self-determination, self-efficacy, intrinsic motivation, and overall science motivation through participating in the CREST programme.

Nye *et al.* (2004) found that teacher effects for promoting academic achievement were lower in higher socio-economic status (SES) schools compared to lower SES schools. As this research was conducted in a rural school in Glasgow, arguably a high SES school, the lack of classroom differences in response to the CREST programme could be explained by considering the findings reported by Nye *et al.* (2004). However, as no data was collected relating to the SES status of the school and the students taking part in the study, further research is needed to assess this explanation. Nye and colleagues (2004) also highlighted that finding no teacher effects does not mean that all teachers demonstrate similar effectiveness in the classroom practice being studied. This highlights the importance of gaining an understanding of teachers' perceptions of the programme studied in this research.

The lack of classroom differences found in the present study may also be explained through considering the research conducted by Skibbe, Phillis, Day, Brophy-Herb, and Connor (2012). Skibbe and colleagues (2012) conducted research in the United States investigating classroom effects in students between the ages of 6 and 10 years. They found that classrooms with stronger self-regulated learners experienced more gain through a reading comprehension and vocabulary skills intervention than those with lower self-regulating peers. These researchers explained that students with lower levels of self-reported regulation may distract their peers, interrupt teaching, and may also require more teacher attention and intervention. Within the context of the research findings presented in this second empirical paper, it is possible that the nine CREST classes had similar distributions of self-regulated learners, and therefore no class differences in response to the intervention were noted. The results from the preliminary analyses documenting that classes were matched on pre-test self-reports provide further support for this explanation.

There is also some research to suggest that classroom and teacher effects occur over time and do not surface immediately (Nye *et al.*, 2004). It is therefore possible that any differences between classes regarding changes in self-reported levels of the measured variables in the present study might develop further after the intervention, and surface at the four-month delayed post-test. This would explain the lack of class differences seen on immediate post-tests. However, as only a subset of students completed the four-month delayed post-test, the present study design was limited to looking at class differences using the pre- and post-test data.

Educational Significance of the Findings

One of the most prominent findings documented in the present paper (and in Author 2016 and Author et al., 2013) was the influence of participation in the CREST programme on students' self-reported levels of self-regulation and self-regulated learning. This work documented significant increases in these constructs from pre-test to post-test for students taking part in the programme that were not observed in the reference control groups. The magnitude of the effects relating to self-regulated processes in the present study were large (as in Author 2016 and Author et al., 2013). Together, these results suggest that giving students opportunities to control and evaluate their learning and work collaboratively with peers toward personally set goals appears to influence their ability to self-regulate their learning in science. It is necessary at this point to translate what these effects mean for teachers and why they are impressive in the context of wider literature in the area.

Considering the above results in the light of downward trends in self-regulation and self-regulated learning over a school term, as reported by Berger and Karabenick (2011), highlights the potential significance of these findings. In more technical terms, the research evidence outlining that smaller effect sizes are documented in controlled studies that use standardised measurement tools underlines the educational significance of these results for teachers and students (Chiu, 1998; Hattie *et al.*, 2009). Finally, as Hattie and colleagues (1996) outlined that medium effect sizes (e.g. d=.4) should be used as a benchmark for discussing research findings as educationally significant, the practical implications of the work presented in this paper are clear. Together with our previous work (Author, 2016; Author et al., 2013), it is clear that the CREST programme has an impact on students' self-regulatory processes that warrant further discussion and study.

Comparing the results presented in this paper to our previous work, larger effect sizes were reported for this sample of students. While these results may be explained by the particular research design of each of the three studies, with the tighter control provided in (Author, 2016) through using two intervention conditions, considering the results in relation to relevant literature may provide further insight. Dignath and Büttner (2008) conducted research investigating the influence of the length of similar interventions in maths and found, unsurprisingly, that interventions with more sessions (ie. more overall time investment) had an increased impact on academic performance in both secondary and primary schools. The larger effect sizes reported in the present study (total of ≅22 hours on CREST) compared to both Author et al. (2013) (≅10 hours) and Author (2016) (≅12 hours) are in agreement with the findings of Dignath and Büttner (2008). In addition, Chiu (1998) found that less intense interventions were more effective (with intensity measured as the average number of days a week spent participating in the programme, ie. fewer days suggesting higher intensity). From this perspective, the larger effect sizes reported in the present study (with 24 sessions over 8 weeks) are not surprising. In addition, as reported above, the guidelines prescribe a minimum of 10 hours for successful completion of the CREST programme. Considering the findings reported here, and remembering our previous work, it is clear that some schools are going above and beyond these requirements. While this is a positive observation, it may warrant further consideration regarding communication of programme guidelines within schools taking part.

The interpretations discussed above provide support for teachers and administrators to conduct the CREST programme over longer periods of time instead of condensing the programme and shortening the implementation period. It is necessary however to understand that this might provide important limitations regarding the nature of the student investigations possible. However, Haller *et al.* (1988) reviewed 20 studies with school children between grades 2 and 12 (aged 6-18) developing metacognitive strategies and found that instruction even as short as 10 minutes per lesson was effective in increasing reading comprehension. Therefore, it may be more effective to have less intense CREST programme work and structure sessions as only a small component of the classroom period over several weeks.

Methodological Limitations and Future Research

A key internal validity issue in the present study is that, due to the lack of ability, SES, and family structure data, there is no way to know how equivalent the classes are on key background variables. While it was stated that groups were matched at pre-test on the variables measured in the study, it is not possible to discern whether the classes were matched on unmeasured pre-test variables (Nye *et al.*, 2004). In addition, no covariates controlling for these pre-existing differences in background variables among students in each classroom could be included in this study. As Rowan, Correnti, and Miller (2002) identified the confounding effects of SES and prior academic performance on classroom differences in achievement, it is possible that these effects extend to the self-regulated processes and related motivations studied here as well. However, while it is unfortunate that previous achievement data could not be obtained, randomisation of students into classes by registration should make pre-test achievement score adjustment unnecessary (Nye *et al.*, 2004).

Considering the ANOVA analyses and paired-samples results presented earlier allows several insights to be gained. The results highlight the sensitivity of small group sizes in variance analyses and the limitations of paired-samples *t-tests* regarding direct group comparisons. It is possible, in the present study, that the absence of robust differences in the ANOVA analyses presented earlier on the change scores of the nine CREST classes was due to the small number of students in each class (Sun *et al.,* 2010). This issue has been documented as one of the main limitations of classroom effect studies due to the high sample sizes needed and the inherent small sample sizes within natural class structure designs, with classes usually ranging from between 20 and 30 students.

It is important to note that the ANOVA analyses presented in this study could also be masking smaller class differences as only 2 out of the 9 classes showed significant changes through CREST programme participation and therefore, the overall weight of the change may have been pulled down by the other CREST classes. However, as highlighted in the methods outlined for the present study, the sample size was slightly less than anticipated which contributed to lower power than required for the paired-samples *t-tests*. This may explain the lack of significant paired-samples *t-tests* for the CREST classes individually. While the findings presented in the final results section of this study addressing class differences by conducting ANOVAs on the change scores for the nine CREST classes may seem to undercut the findings presented previously from the *t-tests,* the inclusion of these results is very important in order to highlight the limitations regarding investigating classroom effects.

The limitation of the present study regarding retention effects, as well as Author et al. (2013), is that not all students were followed longitudinally. In the first study, only students taking part in the CREST programme were given delayed post-tests and in the present study, this data was available for only a subset of the CREST students. This limits not only the interpretation of results, but also the analyses that could have been performed had all students in the study completed the four-month delayed post-test. These issues were considered when designing our final study in this series (Author 2016).

Another aspect of the present methods that informed the design of Author (2016), related to which differences to investigate. Van Horn *et al.* (2008) state that intervention main effects are important in intervention research but highlight that additional insight can also be gained through evaluating potential variations for subgroups of participants. This was considered and incorporated into the design Author (2016).

As was the case regarding our previous work, it is not possible to unpack which aspects of the CREST programme contributed to the changes seen here on the measured outcome variables. Similar difficulties have been documented in intervention studies (De Corte *et al.,* 2004; Glaser & Brunstein, 2007; Williams & Binnie, 2003) and are considered further in Author (2016). In addition to the limited information regarding which aspects of the programme are influencing self-regulated processes and related motivations in students, it was not possible in the present study to discern classroom effects from teacher effects. More qualitative data may provide a solution for this issue through obtaining the perspectives of teachers and students as to why they experienced what they did through CREST programme participation. Looking at teacher beliefs and their influence on the effectiveness of the CREST programme may also provide additional insight (Moos & Ringdal, 2012; Sugrue, 1997).

Finally, as this study included ten classrooms from one rural school in Glasgow, it is necessary for further research to investigate these group and class differences across several school settings in order to obtain reasonable measurement precision and allow for the findings to be generalisable to other educational settings (Nye *et al.,* 2004). In addition, similar to Author et al. (2013), the gender split of the sample was not equal which could also present limitations in terms of the generalisability of the results presented in this paper. More specifically, the gender make-up of the reference control class, with relatively more boys than in each of the nine CREST classes, may have contributed to the lack of significant pre- to post-test changes seen in this class.

CONCLUSIONS

Building on the literature in the area and our previous work, the present study aimed to investigate the impact of the CREST programme on students' self-reported levels of self-regulated processes and related motivations. This work additionally investigated how different classes of students respond to the CREST programme regarding changes in their self-reported levels of the key self-regulation and motivational constructs measured. The study included a total of 178 students and was conducted in an individual school setting with students from one year-group. Similar to Author (2016)and Authors (2013), previously published and validated measures of self-regulatory processes and related motivations were included as dependent variables.

However, in the present study, student classroom allocation was also used as the independent variable in order to investigate the presence of any differences between the nine CREST classes in terms of changes in self-reported levels on the measured variables.

The results showed that overall, students experienced significant increases in levels of self-reported self-regulated learning, self-determination, self-efficacy, intrinsic motivation and personal relevance, and overall science motivation through participation in the CREST programme. By contrast, a reference control class of students not taking part in the programme showed no significant changes in any measured outcomes. These results provided some support that students participating in an inquiry-based learning programme showed different changes from pre-test to post-test compared to the reference control class and replicated some of the findings of Authors(2013) relating to retention effects. The findings have wider implications for science teachers in that they show positive impact (with large and medium effect sizes) following participation in open-inquiry activities. Comparing the effect sizes to our previous work, these findings also offer several recommendations for practice (ie. less intense sessions conducted over a longer period of time). The findings also support the viability of looking at these inquiry activities through a self-regulatory lens.

However, as comparisons between the 160 CREST students and 18 reference control students were limited, these results were interpreted with caution. The findings documented in this paper also revealed no significant differences between changes in self-reported levels on the outcome variables measured among the nine classes included in this study that participated in the CREST programme. These results highlighted the sensitivity of classroom effect studies regarding the choice of analyses and sample size limitations. Further insight was also provided regarding the sensitivity of the self-report measures used in this research.

REFERENCES

Abdullah, Y., & Lee, M. N. (2007, November). *Exploring children's self-regulated learning skills*. Paper presented at the International Conference of Educational Reform Thailand.

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,* 617–639.

Berger, J. L., & Karabenick, S. A. (2011). Motivation and students’ use of learning strategies: Evidence of unidirectional effects in mathematics classrooms. *Learning and Instruction, 21,* 416–428.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7,* 161-186.

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: An International Review, 54 (2),* 199-231.

Boekaerts, M., & Niemivirta, M. (2000). Self-regulated learning: Finding a balance between learning goals and ego-protective goals. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of Self-Regulation* (p. 417-450). San Diego, CA: Academic Press.

Bonate, P. L. (2000). Analysis of pretest-posttest designs. Chapman & Hall/CRC.

Butler, D. L. (1998). The strategic content learning approach to promoting self-regulated learning: A report of three studies. *Journal of Educational Psychology, 90 (4),* 682-697.

Campbell, D. T., & Stanley, J. C. (1966). Experimental and Quasi-Experimental Designs for Research. Houghton Mifflin Company. London. Chicago, IL: Rand- McNally.

Chase, M. A. (2001). Children’s self-efficacy, motivational intentions, and attributions in physical education and sport. *Research Quarterly for Exercise and Sport, 72,* 47–54.

Chularut, P., & Debacker, T. K. (2004). The influence of concept mapping on achievement, self-regulation, and self-efficacy in students of English as a second language. *Contemporary Educational Psychology, 29*, 248-263.

Cleary, T. J., & Chen, P. P. (2009). Self-regulation, motivation, and math achievement in middle school: Variations across grade level and math context. *Journal of School Psychology*, *47*, 291-314.

Cleary, T. J., & Zimmerman, B. J. (2004). Self-regulation empowerment program: A school-based program to enhance self-regulated and self-motivated cycles of student learning. Physchology in Schools, 41(5), 537-550.

Cumming, G., Fidler, F., Kalinowski, P. and Lai, J. (2012), The statistical recommendations of the American Psychological Association Publication Manual: Effect sizes, confidence intervals, and meta-analysis. *Australian Journal of Psychology, 64,* 138–146.

De Boer, H., Donker-Bergstra, A. S., & Kostons, D. D. N. M. (2012). *Effective strategies for self-regulated learning: A Meta-Analysis.* Groningen: GION. p. 72.

De Corte, E., Verschaffel, L., & Masui, C. (2004). The CLIA-model: A framework for designing powerful learning environments for thinking and problem solving. *European Journal of Psychology of Education, 19,* 365-384.

DfE (2013) National Curriculum for England [electronic version]. Available at: www.education.gov.uk/schools/teachingandlearning/curriculum/secondary/b00198831/ science/ks4.

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 Learning, 3*, 231–264.

Dillon, J. (2008). A Review of the Research on Practical Work in School Science. Technical Report, King's College London.

Dimitrov, D. M., & Rumrill, P. D. (2003). Pretest-posttest designs and measurement of change. *Work, 20,* 159-165.

Driver, R. (1989). Students' conceptions and the learning of science. *International Journal of Science Education, 11,* 481-490.

Driver, R., & Oldham, V. (1986). A constructivist approach to curriculum development in science. *Studies in Science Education, 13,* 105-122.

Duckworth, K., Akerman, R., MacGregor, A., Salter, E., & Vorhaus, J. (2009). Self-regualted learning: a literature review. Centre for Research on the Wider Benefits of Learning, *Research Report 33* (p. 1-98). Leading education and social research, Institude of Eduation, University of London.

Duncan, R. G., & Tseng, K. A. (2010). Designing project-based instruction to foster generative and mechanistic understandings in genetics. *Science Education, 95(1),* 21-56.

Fensham, P. J. (2009). The link between policy and practice in science education: The role of research. *Science Education, 93,* 1076-1095.

Field, A. (2013). *Discovering Statistics Using SPSS (3rd Ed.).* Sage Publications, London.

Field, A., & Hole, G. (2003). *How to design and report experiments.* Sage publications, London.

Fraser, B. J., Giddings, G. J., & McRobbie, C. J. (1995). Evolution and validation of a personal form of an instrument for assessing science laboratory classroom environments. *Journal of Research in Science Teaching,* 32(4), 399-422.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-types domains. *Developmental Psychology, 38,* 519–533.

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., & Schroeter, K (2003). Enhancing third-grade students mathematical problem solving with self-regulated learning strategies. *Journal of Educational Psychology, 95(2),* 306-315.

Gao, Z., Lee, A. M., Solomon, & Zhang, T. (2009). Changes in middle school students' motivation toward physical education over one school year. *Journal of Teaching in Physical Education, 28(4),* 378-399.

Glaser, C., & Brunstein, J. C. (2007). Improving fourth-grade students' composition skills: Effects of strategy instruction and self-regulation procedures. *Journal of Educational Psychology, 99(2),* 297-310.

Glynn, S. M., Taasoobshirazi, G., & Brickamn, P. (2009). Science motivation questionnaire: Construct validation with nonscience majors. *Journal of Research in Science Teaching, 46*(2), 127-146.

Graham, S., & Harris, K. R. (1993). Self-regulated strategy development: Helping students with learning problems develop as writers. *The Elementrary School Journal, 94(2),* 169-181.

Haller, E., Child, D. A., & Walberg, H. J. (1988). Can comprehension be taught? A quantitative synthesis of "metacognitive" studies. *Educational Researcher, 17(8),* 5-8.

Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement.* London & New York: Routledge, Taylor& Francis Group, p. 379.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66(2),* 99-136.

Hartman, H. J. (2001). Teaching metacognitively. In H. J. Hartman (Ed.), *Metacognition in Learning and Instruction: Theory, Research and Practice* (p. 33–68). Boston: Kluwer.

Hodgson, C., & Pyle, K. (2010). A literature review of assessment for learning in science. Slough: NFER. Online.

Kaplan, A. (2008). Clarifying metacognition, self-regulation, and self-regulated learning: What's the purpose? *Educational Psychology Review, 20*, 477-484.

Kistner, S., Rakoczy, K., Otto, B., Dignath-van Ewijk, C., & Büttner, G. (2010). Promotion of self-regulated learning in classrooms: Investigating frequency, quality, and consequences for student performance. *Metacognition Learning, 5,* 157-171.

Kohn, A. (2011). The Case Against Grades. *Educational Leadership, 69(3),* 28-33.

Kramarski, B., & Michalsky, T. (2009). Investigating preservice teachers' professional growth in self-regulated learning environments. *Journal of Educational Psychology, 101*(1), 161-175.

Lombraerts, K., Engels, N., & Anthanasou, J. A. (2007). Development and validation of the self-regulated learning inventory for teachers. *Perspectives in Education, 25(4),* 29-48.

MacLellan, E. S., & Soden, R. (2006). Facilitating self-regulation in higher education through self-report. *Learning Environments Research, 9*, 95-110.

Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101(3),* 689-704.

Mikroyannidis, A., Connolly, T., & Law, E. L. (2012). A survey into the teacher's perception of self-regulated learning. *Advanced Learning Technologies (ICALT), 2012 IEEE 12th International Conference proceedings, 696-697.*

Montalvo, F. T., & Torres, M. C. (2004). Self-regulated learning: Current and future directions. *Journal of Research in Educational Psychololgy, 2*(1), 1-34.

Moos, D. C., & Ringdal, A. (2012). Self-regulated learning in the classroom: A literature review on the teacher's role. *Educational Research International, Volume 2012, Article ID 423284,* p. 15.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26(3),* 237-257.

Perels, F., Gurtler, T., & Schmitz, B. (2005). Training of self-regulatory and problem-solving competence. *Learning and Instruction, 15(2),* 123-139.

Orrow-Whiting, M., Edwards, R. and Slade, P. (2007) Keeping up the momentum: modernising the secondary science curriculum in England. *School Science Review, 89(326)*, 27–29.

Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of providing choices in the classroom. *Journal of Educational Psychology, 102*, 896-915.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95(4),* 667-686.

Pintrich, P. R., & De Groot, E. V. (1990). Motivation and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33-40.

Randi, J., & Corno, L. (1997). Teachers as innovators. In B.J. Biddle, T.L. Good, & I.F. Goodson (Eds.), *International handbook of teachers and teaching, Vol. I*. (pp. 1163–1221). Dordrecht, The Netherlands: Kluwer Academic.

Reif, F. (2008). *Applying cognitive science to education: Thinking and learning in scientific and other complex domains.* Cambridge, MA: MIT Press.

Rowan, B., Correnti, R., & Miller, R. (2002). What large-scale, survey research tells us about teacher effects in student achievement: Insights from the Prospects Study of elementary schools. Madison, WI: Consortium for Policy Research in Education. Retrieved on June 25th 2017, from: http://www.cpre.org/images/stories/cpre_pdfs/rr51.pdf

Rozendaal, J. S., Minnaert, A., & Boekaerts, M. (2005). The influence of teacher perceived administration of self-regualted learning on students' motivation and information-processing. *Learning and Instruction, 15,* 141-160.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25,* 54-67.

Ryder, J. (2002). School science education for citizenship: Strategies for teaching about the epistemology of science. *Journal of Curriculum Studies, 35 (6),* 637-658.

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*, 111-139.

Schunk, D. H., & Ertmer, P. A. (2000). Self-regulation and academic learning: Self- efficacy enhancing interventions. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (p. 631–649). New York: Academic.

Severiens, S., Ten Dam, G., & Van Hout Wolters, B. (2001). Stability of processing and regulation strategies: Two longitudinal studies on student learning. *Higher Education, 42,* 437-453.

Skibbe, L. E., Phillis, B. M., Day, S. L., Brophy-Herb, H. E., & Connor, C. M. (2012). Children's early literacy growth in relation to classmates' self-regulation. *Journal of Educational Psychology, 104(3),* 541-553.

Smolkowski, K. (2013). Gain score analysis. Retrieved on September 7th, 2017 from: http://homes.ori.org/keiths/Tips/Stats_GainScores.html

Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology, 102(4),* 947-963.

Stoeger, H., & Ziegler, A. (2010). Do pupils with differing cognitive abilities benefit similarly from a self-regualted learning training program? *Gifted Education International, 26(1),* 110-123.

Sun, S., Pan, W., & Wang, L. L. (2010). A comprehensive review of effect size reporting and interpreting practices in academic journals in education and psychology. *Journal of Educational Psychology, 102(4),* 989-1004.

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Science, 312*, 1143-1144.

Throndsen, I. (2011). Self-regulated learning of basic arithmetic skills: A longitudinal study. *British Journal of Educational Psychology, 81,* 558-578.

Van der Veen, I., & Peetsma, T. (2009). The development in self-regulated learning behavior of first-year students in the lowest level of secondary school in the Netherlands. *Learning and Individual Differences, 19(1),* 34-46.

Van Horn, M. L , Fagan, A. A., Jaki, T., Brown, E. C., Hawkins, D., Arthur, M. W *et al.* (2008). Using multilevel mixtures to evaluate intervention effects in group randomized trials. *Multivariate Behavioral Research, 43,* 289-326.

Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K., & Lens, W. (2009). Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of Educational Psychology, 101*(3), 671-688.

Wasserstein, R.L., & Lazar, N. (2016). The ASA's statement on p-Values. Context, process, and purpose. *The American Statistician, 70(2),* 129-133.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16(1),* 3-118.

Wigfield, A., & Eccles, J. S. (2002). Students' Motivation During the Middle School Years. In J. Aronson (Ed.), *Improving Academic Achievement: Impact of Psychological Factors on Education* (p. 395). London: Academic Press.

Williams, J. M, & Binnie, L. M. (2003). Children's concepts of illness: An intervention to improve knowledge. *British Journal of Health Pscyhology, 7(2),* 129-147.

Winne, P. H. (1995). Self-regulation is ubiquitous but its forms vary with knowledge. *Educational Psychologist, 30(4),* 223-228.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future projects. *American Educational Research Journal, 45*(1), 166-183.

Zimmerman, B. J. (2000). Attaining self-regulated learning: A social cognitive persepctive. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self- regulation* (p. 13-39). San Diego, CA: Academic Press.

Zimmerman, D. W. (1997). A note on interpretation of the paired-samples t test. *Journal of Education and Behavioral Statistics, 22(3),* 349-360.

Zimmerman, B. J., & Martinez-Pons, M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology, 80*(3), 284-290.

Zimmerman, D. W, & Zumbo, B. D. (1993). The relative power of parametric and nonparametric statistical methods. In Karen, G., & Lewis, C. (Eds.), *A Handbook on Data Analysis in the Behavioral sciences: Statistical Issues* (p. 481-517). *Lawrance Erlbaum Associates; Hilsdale, New Jersey.*

Zohar, A. (2004). *Higher order thinking in science classrooms: Students' learning and teacher' professional development*. The Netherlands: Kluwer Academic Press.

Zohar, A., & Dori, Y. (Eds.). (2012*). Metacognition in Science Education: Trends in Current Research. London, UK: Springer.*

Zusho, A., Pintrich, P. R., & Coppola, B. (2003). Skill and will: The role of motivation and cognition in the learning of college chemistry. *International Journal of Science Education, 25(9),* 1081-1094.

Notes

[i] The CREST programme can be embedded into natural classrooms and used as a tool to work towards several key curriculum objectives while also developing these regulatory and motivational skills. The structure of the programme (eg. how many sessions over what time interval) is flexible in order to fit with school and teacher activities. Students work through projects, with support to guide them, and

awards are made at three levels depending on the time commitment for the project: Bronze (10 hours), Silver (30 hours), and Gold (70 hours). This study investigates the Broze level of the award programme.