



What predicts performance during clinical psychology training?

Katrina Scior^{1*}, Caroline E. Bradley¹, Henry W. W. Potts²,
Katherine Woolf³ and Amanda C de C Williams¹

¹Research Department of Clinical, Educational and Health Psychology, University College London, UK

²CHIME (Centre for Health Informatics & Multiprofessional Education), Institute of Epidemiology & Health Care, University College London, UK

³UCL Medical School, University College London, UK

Objectives. While the question of who is likely to be selected for clinical psychology training has been studied, evidence on performance during training is scant. This study explored data from seven consecutive intakes of the UK's largest clinical psychology training course, aiming to identify what factors predict better or poorer outcomes.

Design. Longitudinal cross-sectional study using prospective and retrospective data.

Method. Characteristics at application were analysed in relation to a range of in-course assessments for 274 trainee clinical psychologists who had completed or were in the final stage of their training.

Results. Trainees were diverse in age, pre-training experience, and academic performance at A-level (advanced level certificate required for university admission), but not in gender or ethnicity. Failure rates across the three performance domains (academic, clinical, research) were very low, suggesting that selection was successful in screening out less suitable candidates. Key predictors of good performance on the course were better A-levels and better degree class. Non-white students performed less well on two outcomes. Type and extent of pre-training clinical experience on outcomes had varied effects on outcome. Research supervisor ratings emerged as global indicators and predicted nearly all outcomes, but may have been biased as they were retrospective. Referee ratings predicted only one of the seven outcomes examined, and interview ratings predicted none of the outcomes.

Conclusions. Predicting who will do well or poorly in clinical psychology training is complex. Interview and referee ratings may well be successful in screening out unsuitable candidates, but appear to be a poor guide to performance on the course.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

*Correspondence should be addressed to Dr Katrina Scior, Research Department of Clinical, Educational & Health Psychology, University College London, Gower Street, London WC1E 6BT, UK (email: k.scior@ucl.ac.uk).

Practitioner points

- While referee and selection interview ratings did not predict performance during training, they may be useful in screening out unsuitable candidates at the application stage
- High school final academic performance was the best predictor of good performance during clinical psychology training
- The findings are derived from seven cohorts of one training course, the UK's largest; they cannot be assumed to generalize to all training courses

Clinical psychology training in the United Kingdom is solely offered on a post-graduate basis, with a recognized degree in psychology as a prerequisite. It involves training in three areas over the course of three calendar years: academic teaching and study, clinical placements in the National Health Service (NHS), and research including completion of a doctoral thesis. Successful completion of all aspects leads to the award of a Doctorate in Clinical Psychology (DClinPsy or ClinPsyD). Training is underpinned by a scientist-practitioner model similar to the Boulder model adopted in the United States more than 60 years ago (McFall, 2006). Training places for citizens of the United Kingdom or other countries in the European Economic Area are fully funded and salaried by the NHS. All 30 training courses are accredited by the British Psychological Society, and also approved by the Health & Care Professions Council.

Entry to clinical psychology training is highly competitive in the United Kingdom. All applications are administered by a national Clearing House (<http://www.leeds.ac.uk/chpccp>), using a single application for up to four courses across the United Kingdom. The high ratio of applicants to training places (on average 3.8:1 across all UK courses during the period covered by this study, data provided by the Clearing House), and the generally high quality of applications, makes selection resource intensive. In response, there has been a lot of interest over recent years in the selection process. Some courses have introduced computerized or written tests as part of their short-listing process and work is underway to develop a national screening test. Despite their somewhat different selection criteria and processes, all courses shortlist on the application form submitted via the Clearing House and interview as part of the selection process. Although several studies have examined factors associated with application success (Phillips, Hatton, & Gray, 2004; Scior, Gray, Halsey, & Roth, 2007), and selection procedures (Hemmings & Simpson, 2008; Simpson, Hemmings, Daiches, & Amor, 2010), the only UK-based study on prediction of performance during training demonstrated agreement between performance on a written short-listing task and academic performance on the course (Hemmings & Simpson, 2008), albeit with a small sample ($N = 45$). Several US studies have identified differences in the intakes and subsequent career paths associated with the three main clinical psychology training models used in the United States (clinical scientist, scientist-practitioner or practitioner-scholar; Cherry, Messenger, & Jacoby, 2000; McFall, 2006). However, our searches did not reveal any English language studies other than Hemmings and Simpson's (2008) that examined factors associated with performance during training. Given the high resource costs involved in training clinical psychologists, and the substantial responsibility and power trainees have on qualification, it is surprising that evidence on predicting good or poor performance during clinical psychology training is sparse. A likely reason for this gap is the fact that attrition and failure are rare in clinical psychology training, requiring large samples to investigate. Furthermore, courses vary somewhat in assessment procedures, making it difficult to assess outcomes across training courses.

Predictors of underperformance in medicine

It is useful to refer to the extensive literature on the predictors of dropout and academic underperformance at medical school. While medicine and clinical psychology have many differences, they also have important similarities: both require academic proficiency combined with communication and professional skills; both are highly selective; both qualify UK graduates for employment in the NHS; both have a strong professional identity.

Medical admissions procedures attempt to select for academic and non-academic competence using a combination of school grades, aptitude test scores, personal or school statements, and interviews (Parry *et al.*, 2006). As in clinical psychology, there is concern in medicine about the demographic profile of the student population, especially in terms of gender and socioeconomic background (Elston, 2009; Mathers, Sitch, Marsh, & Parry, 2011). The medical education literature therefore explores which factors predict medical school outcomes.

Pre-admission grades

Pre-admission grades are consistent predictors of medical school performance (Ferguson, James, & Madeley, 2002). Lower school grades also predict dropout (O'Neill, Wallstedt, Eika, & Hartvigsen, 2011). However, the majority of medical applicants have top grades (McManus *et al.*, 2005), leading to the use of aptitude tests in selection, although with much debate about their usefulness (Emery, Bell, & Vidal Rodeiro, 2011; McManus *et al.*, 2005). Tests for selection into undergraduate medical courses in use in the United Kingdom and Australia seem not to have good predictive validity (McManus, Ferguson, Wakeford, Powis, & James, 2011; Mercer & Puddey, 2011; Wilkinson *et al.*, 2008; Yates & James, 2010), but MCAT, the test used in the United States where medicine is a graduate course like clinical psychology, appears to have reasonable predictive power (Donnon, Paolucci, & Violato, 2007).

Interviews

Traditional interviews have low predictive power (Goho & Blackman, 2006), and variations in interviewing methods make it hard to identify consistent relationships between interview characteristics and outcomes (Ferguson *et al.*, 2002). Many medical schools now use the multiple-mini interview in which students are assessed on how they deal with professional situations in practice (Eva *et al.*, 2009). This seems to predict performance on later similar practical tests at medical school (Eva, Rosenfeld, Reiter, & Norman, 2004).

Personal and academic references

References are generally poor predictors (Ferguson, James, O'Hehir, Sanders, & McManus, 2003; Siu & Reiter, 2009), although negative comments from an academic referee may predict in-course difficulties (Yates & James, 2006), and a Canadian study showed personal statements and references to have a small predictive effect on medical school clinical performance (Peskun, Detsky, & Shandling, 2007).

In the United Kingdom, medical school performance is also associated with demographics, with females and white students doing better, raising issues of equity (Ferguson *et al.*, 2002; Higham & Steer, 2004; Woolf, McManus, Potts, & Dacre, 2013; Woolf, Potts, & McManus, 2011).

Range restriction and other statistical challenges

Studies generally correlate selection variables (e.g., pre-admission grades, interview performance) with outcome measures. However, outcome measures are only available on those who were selected and not, naturally, on those who were not accepted. How selection variables predict performance within the selected group is not the same as assessing the variables as a means of selection (Burt, 1943). Because of range restriction, observed correlations between selection variables and outcomes are generally smaller than they would be were outcomes available on all applicants. It is possible to make a correction for range restriction, but only with the right data available on non-successful applicants (Sackett & Yang, 2000).

Other issues in the data – like the reliability of measures, ceiling effects, and ordinal outcome data – may also have the effect of reducing the observed correlations. Because of all these factors, the construct-level predictive validity can be much higher than observed correlations between selection variables and course outcomes (McManus *et al.*, in press).

This study

The clinical psychology doctorate course investigated is the largest one in the United Kingdom, with an average applicant to place ratio of around 28:1 for its 40–42 training places. The course employs a three-stage selection procedure involving course staff and local clinical psychology supervisors. Written guidelines for selectors aim for maximum fairness in selection. The course previously found that successful applications were predicted by A-level (academic qualification offered by educational institutions in England, Wales, and Northern Ireland to students completing high school education) points (see Methods) and academic and clinical referee ratings (Scior *et al.*, 2007), although selectors may rely particularly on these in the absence of other clear ways of distinguishing among hundreds of applicants with good honours degrees.

The aim of this study was to investigate the role of applicant characteristics, interview ratings, and referee ratings in predicting course performance in three domains: academic, clinical, and research. In doing so, we aimed to inform future selection procedures by identifying the predictive power of information available to selectors.

Methods

Participants

Overall, 274 trainee clinical psychologists (the entire 2002–2008 entry cohorts) who had completed all aspects of their training by the time of the study, or were in the process of making revisions to their doctoral theses, were included in the study (as was one individual who had completed all aspects of training other than the thesis due to an extension). Over the seven cohorts studied, the annual intake increased from 30 to 42. Two trainees dropped out of training during the period studied, both within the first year of training, due to personal reasons. Due to the small numbers concerned, it was not feasible to examine what factors may influence dropout and these two were not included in the study.

Application form data

Information about demographics, educational, and employment histories was taken from each Clearing House application form. A-level points were calculated using the British

Universities and Colleges Admissions Service tariff points system (grade A = 120 points; B = 100; C = 80; D = 60; E = 40), with points for each A-level subject added to make up a composite score. The degree class awarded for the first university degree was recorded, even when not psychology. For applicants with a 2.1 and a percentage mark, the degree was classified into low 2.1 <63.9%, mid 2.1 64–66.9%, or high 2.1 >67%. Whether the trainee had completed an MSc and/or PhD was recorded as binary data. The type of school attended was classified into state, grammar, or independent. The type of university (for first degree) was classified into Oxbridge, Russell Group, other pre-1992 universities, post-1992 universities, and non-UK universities.

A substantive voluntary post in the NHS and/or work as a Research Assistant was recorded separately. Overall, clinical experience was classified as minimal (<1 year or mix of short-term voluntary and paid part-time roles, or roles not highly relevant to clinical psychology), moderate (at least 1 year of a role highly relevant to clinical psychology, for example, assistant psychologist, graduate mental health worker, low intensity Improving Access to Psychological Therapy (IAPT) programme worker, research assistant on a clinical project), or substantial (work at the 'moderate' level in a range of roles and services). Ratings were made by the first and last authors independently after rating several application forms together and discussing them to achieve consistency.

Selection interview ratings

We used ratings recorded from interviews and held in the database. Interview panels consisting of three interviewers, at least one member of course staff and at least one regional supervisor rated each interviewee on a 10-point scale, with 10 denoting *outstanding performance* and 0 *exceptionally poor performance* and all interim scale points having clear descriptors. From 2002 to 2006, academic (A) and clinical (B) interview performance was rated; from 2007 onwards, a third category of overall personal suitability was added (C), referring to communication skills, interpersonal style, reflexivity and self-awareness, and readiness to train. While service users advise the course on its selection procedures and interview questions, they are not directly involved in interviews, not least due to the practical challenges arising from holding large numbers of interviews, so it was not possible to obtain service user ratings.

Outcome data

These included contemporary outcomes, measured at the time of assessment, and retrospective ratings gathered for this study.

Contemporary outcomes

Academic performance. Case report scores and exam marks. Each trainee completed five case reports, marked as 'pass', 'minor revisions', 'stipulated revisions', 'major revisions', or 'fail'. An overall case report score was calculated as the number of case reports that were given stipulated or major revisions or failed. Trainees took two exams in year 1 and two in year 2, and scores were z-transformed to account for differences across cohorts. Although marks are analysed, trainees were only required to pass the exam to progress.

Clinical performance. Number of major concerns about performance on placement reported to exam boards, and number of placement failures.

Research ability. Clinical viva result with 'pass' and 'one-month corrections' combined into one category, and compared to 'three-month corrections', 'one-year corrections' or 'fail'.

Retrospective ratings

Course tutors were asked (yes/no) whether they had global concerns about their former tutees in three areas: interpersonal skills (A), robustness (B), and ability to think critically (C).

Research supervisors retrospectively rated trainees' research performance using a 5-point scale anchored with 0 = *poor*; 3 = *average*; 5 = *outstanding*. Supervisors were asked to consider six factors when making one overall rating: (1) the trainee's capacity to think scientifically; (2) research analysis skills; (3) quality of written work; (4) critical thinking skills; (5) organization and planning abilities; and (6) ability to work autonomously.

Results

Descriptive statistics

Demographics

The mean age of trainees at entry to the course was 27 years (range 21–51 years). Eighty-five percentage were females ($n = 234$), 15% were males ($n = 40$); 9% were from black and minority ethnic (BME) backgrounds ($n = 25$). On these indicators, trainees were very similar to the equivalent training cohorts across the United Kingdom of whom 85.3% were females and 8.9% were from BME backgrounds (Clearing House data). Application form data on nine trainees were missing.

Prior qualifications and experience

There was marked variation in trainees' A-level performance. The mean A-level composite score was 353 (range 160–480). One hundred and thirty-one trainees had attended a state school, 42 a grammar school and 85 an independent school; data were unobtainable for 16. 40% had a first class degree ($n = 109$), 55% a 2.1 ($n = 150$), and 2% a 2.2 ($n = 5$). Those with a 2.2 degree had subsequently shown strong performance in a subsequent undergraduate or post-graduate degree. Information about the type of degree could only be ascertained for half of the trainees with a 2.1 ($n = 75$). Of these, 71% had a high 2.1 ($n = 53$), 16% a mid-2.1 ($n = 18$), and 13% a low 2.1 ($n = 10$). Thirty-five trainees had attended Oxford or Cambridge Universities, 27 Russell Group Universities, 156 other pre-1992 universities, 35 post-1992 universities, and 11 non-UK universities (six of which were in the Republic of Ireland); data were unobtainable for 10. Prior qualifications were similar when comparing white and BME trainees. The two groups had similar A level scores (composite score for white trainees: $M = 354.55$, $SD = 72.33$; for BME trainees $M = 333.63$, $SD = 72.08$), $t(251) = 1.30$, $p = .20$. The two groups did not differ by degree class, $\chi^2(2) = 1.05$, $p = .59$. Of white trainees, 24% had attended Oxbridge or Russell

Group universities and a further 57% other pre-1992 universities, compared to 23% and 59%, respectively, of BME trainees.

The mean time between trainees obtaining their Graduate Basis for Chartering (GBC) with the British Psychological Society (BPS), either as a result of completing an accredited psychology degree or conversion diploma, and beginning the course was 3.3 years ($SD = 2.0$, range 0–16 years). A PhD had been obtained by 5% ($n = 15$), an MSc by 29% ($n = 80$).

Applicants had varied relevant clinical experience at the time of application: 42% had minimal ($n = 116$), 40% moderate ($n = 110$), and 14% substantial ($n = 39$) experience. Eighty-six percentage had worked in the NHS ($n = 237$) and 64% as research assistants ($n = 176$).

References

All candidates were required to submit an academic and a clinical reference at application. Referees provide a rating alongside their narrative reference comparing the candidate to other clinical psychology applicants they had provided references for, using a 5-point scale with two anchors (1 = *much worse than others*; 5 = *the best*). The large number ($n = 66$) of missing ratings is due to referees not having acted as referee previously and thus being unable to compare. The mean rating by both academic and clinical referees was 4.5 (both $SD = 0.6$); no trainee was given a rating below three.

Interviews

Mean interview ratings were: for part A (academic/theory) = 8.3 ($SD = 0.8$), for part B (clinical) = 8.4 ($SD = 0.8$), and for part C (personal suitability) = 8.9 ($SD = 0.8$) (the C rating was only available for the 84 trainees selected from 2007 onwards).

Outcome data

Contemporary outcomes

Academic performance. The median number of case reports marked 'stipulated' or worse was two (interquartile range 1–3; range 0–5). Fifty-eight trainees received major revisions (2 months) or a fail for at least one report, including 11 who failed at least one case report.

Clinical performance. Twelve trainees failed a placement or provoked serious concerns about their performance on placement.

Research performance. Twenty-one trainees received either 3-month ($n = 20$) or 1-year corrections to their theses; none failed their viva. These were combined for further analysis and compared to trainees who received a pass or minor revisions in their viva.

Interrelationships between contemporary outcomes. The relationships between contemporary academic outcomes are shown in Table 1. All exam marks were positively correlated. Higher exam marks were associated with a decreased chance of getting

Table 1. Relationships between contemporary academic outcomes assessed by non-parametric correlations (Kendall's τ_b)

Variable	Exam 1 (psychological theory, year 1)	Exam 2 (research methods, year 1)	Exam 3 (advanced psychological theory, year 2)	Exam 4 (statistics, year 2)
Case report score (number marked stipulated revisions or worse)	$\tau_b = -0.16$; $p = .001$	$\tau_b = -0.21$; $p < .001$	$\tau_b = -0.15$; $p = .001$	$\tau_b = -0.23$; $p < .001$
Exam 1		$\tau_b = 0.31$; $p < .001$	$\tau_b = 0.25$; $p < .001$	$\tau_b = 0.22$; $p < .001$
Exam 2			$\tau_b = 0.22$; $p < .001$	$\tau_b = 0.31$; $p < .001$
Exam 3				$\tau_b = 0.19$; $p < .001$

stipulated revisions or worse on case reports. A Kendall's coefficient of concordance across all four exam marks and the case report score was calculated. This was statistically significant, $W = 0.22$, $p < .001$, confirming that the variables were related to each other, but at quite a low level, so they were analysed separately rather than combining them into a global performance measure. The relationships between contemporary academic and research outcomes are shown in Table 2.

Poor placement performance was related to poor exam performance, but there was no association between poor placement performance and case report marks, which raises the question of the extent to which case reports measure clinical knowledge and/or skills. While 58 trainees received at least one major revision or fail on their case reports, of the 12 trainees with poor performance on placement, only three were among these 58. Trainees who received 3-month or 1-year corrections on their thesis were more likely to have received poorer grades in all exams, and in their case reports.

Retrospective outcomes

Course tutor ratings. Tutors raised concerns about the interpersonal skills of 20 trainees across all intakes; about the robustness of 18 trainees; and the critical thinking

Table 2. Relationships between contemporary academic and research outcomes

Binary variable	Continuous variable	Test result	p
Placement concerns	Exam 1 (Theory)	$t_{272} = 2.3$.025
	Exam 2 (Research)	$t_{272} = 1.5$.13
	Exam 3 (Theory)	$t_{272} = 3.9$	<.001
	Exam 4 (Stats)	$t_{272} = 3.6$	<.001
	Case report score	$U = 1449$.64
Viva outcome	Exam 1	$t_{272} = 1.9$.055
	Exam 2	$t_{272} = 3.7$	<.001
	Exam 3	$t_{272} = 2.3$.025
	Exam 4	$t_{272} = 3.2$.002
	Case report score	$U = 1165.5$	<.001

ability of 18 trainees. Concerns about interpersonal skills were statistically significantly related to concerns about robustness (Fisher exact $p < .001$) and critical thinking (Fisher exact $p = .001$), but the latter two were not significantly related (Fisher exact $p = .11$).

Research supervisor ratings. The mean supervisor rating of trainees' research skills was 3.5 ($SD = 1.0$). Raters used the full 5-point scale: 4% of the sample was rated as showing poor research skills (rating < 2), 13% as showing poor or below average skills (rating < 3), and 16% as outstanding (rating = 5).

Predictors of performance on the DClinPsy course

Multivariate statistics were used generally, but given the small numbers with placement concerns and poor thesis performance, these were analysed using univariate tests only.

Predicting poor clinical performance

Table 3 shows the predictors of placement concerns/failure. A-level points, course tutor concerns in all areas, and research supervisor ratings were associated with poor performance.

Predicting poor research performance (viva outcome)

The predictors of research performance are shown in Table 4. Poorer viva outcome (3-month or 1-year corrections) was associated with course tutor concern over critical thinking, research supervisor ratings, and to a lesser extent with a longer time between obtaining GBC and start of training.

Predicting exam performance

Multiple regression was used to examine the predictors of exam performance. For each exam, an initial model regressed exam performance on to the retrospective ratings. A second model regressed exam performance on to the following pre-course variables: gender, ethnicity, age, A-level points, school type, degree class, university type, time since obtaining GBC, research assistant experience, clinical experience, NHS experience, and whether obtained MSc or PhD. To this, second model were added referee ratings, and then to these, interview ratings were added.

Exam 1 – psychological theory, year 1. The retrospective ratings model was statistically significant, $F_{5,256} = 5.1$, $p < .001$, adjusted $R^2 = 7.3\%$. Only higher research supervisor ratings were significantly associated with better psychological theory exam marks, see Table 5. The pre-course variables regression model was also statistically significant, $F_{17,224} = 2.8$, $p < .001$, adjusted $R^2 = 11.4\%$, with better degree class, higher A-level points, and not having attended a grammar school associated with higher exam 1 marks, see Table 5. Adding referee ratings to the pre-course variables resulted in a significant change in R^2 : $F_{2,141} = 3.7$, $p = .028$, new adjusted $R^2 = 12.2\%$. In this model, only higher degree class significantly positively predicted exam 1 marks. Better referee ratings were associated with *lower* exam 1 marks, but neither rating is quite statistically significant on

Table 3. Univariate predictors of concerns about placement performance/failure

Variable	Test statistic	<i>p</i>
Retrospective Ratings		
Course tutor concern over interpersonal skills	Fisher exact test	.001***
Course tutor concern over robustness	Fisher exact test	.037*
Course tutor concern over critical thinking	Fisher exact test	<.001***
Research supervisor rating	$t_{262} = 3.9$	<.001***
Pre-course variables		
	Gender Fisher exact test	.4
	Ethnicity Fisher exact test	.090
	Age $U = 1783.5$.3
	A-level points $U = 801$.009***
	Degree class $U = 1618$.6
	Time lag since obtaining $U = 1478.5$	1.0
	GBC	
	Research assistant experience Fisher exact test	1.0
	Clinical experience $U = 1900.5$.1
	NHS experience Fisher exact test	.4
	Completed MSc Fisher exact test	.2
	Completed PhD Fisher exact test	1.0
	School type Fisher exact test	.8
	University type Fisher exact test	.3
Referee Ratings		
	Academic reference $U = 998$.5
	Clinical reference $U = 520.5$.06
Interview ratings		
	Academic rating (A) $U = 1507.5$.8
	Clinical rating (B) $U = 1342$.4

Note. MHS = national health service; GBC = graduate basis for chartering.

* $p < .05$; ** $p < .01$; *** $p < .001$

its own ($0.1 > ps > 0.05$). Finally, adding the interview ratings did not result in a significant change in R^2 : $F_{2,139} = 2.9$, $p = .056$. Due to the large number of predictors entered, only those that emerged as significant are shown in Tables 5 and following; full details available on request.

Exam 2 – research methods, year 1. The retrospective ratings model was statistically significant, $F_{5,256} = 10.0$, $p < .001$, adjusted $R^2 = 14.7\%$. Course tutor concerns over interpersonal skills, research supervisor ratings, and to a lesser extent tutor concerns over critical thinking were independently associated with research methods exam performance, see Table 6. The pre-course variables model was also statistically significant, $F_{17,224} = 4.0$, $p < .001$, adjusted $R^2 = 17.5\%$, with white ethnicity, better degree class, better A-levels and not attending a post-1992 or non-UK university independently associated with higher exam 2 marks. Adding the referee ratings did not change the R^2

Table 4. Univariate predictors of concerns about research performance as measured by viva outcome

Variable	Test statistic	<i>p</i>
Retrospective ratings		
Course tutor concern over interpersonal skills	Fisher exact test	.6
Course tutor concern over robustness	Fisher exact test	.4
Course tutor concern over critical thinking	Fisher exact test	.006**
Research supervisor rating	$t_{262} = 3.1$.002**
Pre-course variables		
Gender	Fisher exact test	.7
Ethnicity	Fisher exact test	.4
Age	$U = 2442.5$.7
A-level points	$U = 2166.5$.9
Degree class	$U = 2866.5$.3
Time lag since obtaining GBC	$U = 3118.5$.046*†
Research assistant experience	Fisher exact test	.8
Clinical experience	$U = 3101$.08
NHS experience	Fisher exact test	.7
Completed MSc	Fisher exact test	.3
Completed PhD	Fisher exact test	.8
School type	Fisher exact test	.6
University type	Fisher exact test	.2
Referee Ratings		
Academic reference	$U = 1465$.6
Clinical reference	$U = 1486.5$.9
Interview ratings		
Academic rating (A)	$U = 2155.5$.1
Clinical rating (B)	$U = 2119.5$.1

Note. NHS = National Health Service; GBC = Graduate Basis for Chartering.

* $p < .05$; ** $p < .01$; *** $p < .001$.

†Median time lag for good performance 3 years; for poor performance median time lag 4 years.

Table 5. Multivariate predictors of exam I marks

Variable	<i>B</i>	<i>p</i>
Retrospective ratings		
Research supervisor rating	0.21	.001***
Pre-course variables		
A-level points	0.002	.015*
Degree class	0.41	.001**
Attending a grammar school	-0.39	.035*
Referee ratings added		
Degree class	-0.51	.002**

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 6. Multivariate predictors of exam 2 marks

Variable	B	p
Retrospective ratings		
Course tutor concern over interpersonal skills	-0.78	.002**
Course tutor concern over critical thinking	-0.61	.021*
Research supervisor rating	0.26	<.001***
Pre-course variables		
Ethnicity (being non-white)	-0.65	.003**
A-level points	0.002	.021*
Degree class	0.45	<.001**
Attending a post-1992 university	-0.41	.028*
Attending a non-UK university	-0.76	.029*

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 7. Multivariate predictors of exam 3 marks

Variable	B	p
Retrospective ratings		
Research supervisor rating	0.16	.01*
Pre-course variables		
A-level points	0.009	.007**

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

significantly: $F_{2,141} = 0.4$, $p = .7$, or did subsequently adding the interview ratings: $F_{2,139} = 1.1$, $p = .3$. Referee ratings and interview scores did not predict exam 2 marks.

Exam 3 – advanced psychological theory, year 2. The retrospective ratings model was statistically significant: $F_{5,256} = 4.7$, $p < .001$, adjusted $R^2 = 6.5\%$, and only the research supervisor rating was significantly associated with advanced psychological theory exam performance, see Table 7. The pre-course variables model was also statistically significant, $F_{17,224} = 1.8$, $p = .025$, adjusted $R^2 = 5.6\%$ with only A-level points predicting exam 3 marks. Adding the referee ratings did not change the R^2 significantly: $F_{2,141} = 0.7$, $p = 0.5$, or did subsequently adding the interview ratings: $F_{2,139} = 1.4$, $p = .3$. Referee ratings and interview scores did not predict exam 3 marks.

Exam 4 – statistics, year 2. The retrospective ratings regression model was statistically significant, $F_{5,256} = 7.7$, $p < .001$, adjusted $R^2 = 11.4\%$, and only research supervisor ratings were significantly associated with statistics exam performance, see Table 8. The pre-course variables model was also statistically significant, $F_{17,224} = 4.9$, $p < .001$, adjusted $R^2 = 21.7\%$. Younger age, being white, better A-levels, attending Oxbridge, not attending a post-1992 or non-UK university and *less* clinical experience predicted higher statistics exam marks. Adding referee ratings did not change the R^2 significantly:

Table 8. Multivariate predictors of exam 4 marks (statistics)

Variable	B	p
Retrospective ratings		
Research supervisor rating	0.33	<.001***
Pre-course variables		
Age	-0.39	.033*
Ethnicity (being non-white)	-5.0	.039*
A-level points	0.022	.036*
Attending Oxbridge	4.5	.040*
Attending a post-1992 university	-4.3	.035*
Attending a non-UK university	-8.3	.030*
Clinical experience	-0.21	.034*

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 9. Multivariate predictors of case report marks

Variable	IRR	p
Retrospective ratings		
Research supervisor rating	0.8	<.001***

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

$F_{2,141} = 0.6, p = .6$. Then, adding the interview ratings did not change the R^2 significantly: $F_{2,139} = 0.3, p = .7$. Referee ratings and interview scores did not predict exam 4 marks.

Predicting case report marks

Multiple Poisson regression was performed to predict the number of stipulated, major, and fail marks trainees received for their five case reports, along the same lines as the multiple linear regressions used to predict the exam marks. The retrospective ratings model was statistically significant: $\chi^2(5) = 19.9, p = .001$. Only research supervisor ratings predicted case report marks, see Table 9. A multiple regression using the pre-course variables was not statistically significant: $\chi^2(17) = 17.6, p = .4$. Adding the referee ratings did not change the deviance significantly: $\chi^2(2) = 0.7, p = .7$. Then, adding the interview ratings also did not change the deviance significantly: $\chi^2(2) = 2.1, p = .4$. Referee ratings and interview scores did not predict case report marks.

Poor performance, interview ratings, and pre-training background

Given that interview ratings are a cornerstone of the selection process, we took a closer look at interview ratings for those trainees whose performance in at least one area during training was markedly poor. The 12 trainees with placement concerns were rated somewhat lower in sections A and B of their interviews (section C ratings were omitted as they were only available for one), but none of these differences approached significance. In terms of their clinical experience prior to starting the course, 10 of the 12 had worked in the NHS, four had minimal, three moderate, and five substantial clinical experience.

The data on the 11 trainees who failed at least one case report (one of whom failed two reports) again showed no marked differences on interview ratings, although their mean scores in B and C were somewhat lower than those without failed case reports. Four of these 11 had a first class degree, three an MSc, and one a PhD. Their A-level points ranged from 260 to 480, and three had at least three A-levels at grade A.

The 11 trainees rated as showing poor research skills by their supervisors (rating < 2) received similar ratings on interview parts B and C, but were rated as poorer on section A ($M = 7.9$, $SD = 0.9$) than those not rated poorly ($M = 8.3$, $SD = 0.8$). Of these 11, six had an MSc (again following a 2.1 degree), two a first class degrees, and one a PhD (following a 2.1 degree). Their A-level points ranged from 240 to 360.

Discussion

Completion rates in clinical psychology training are very high, with dropout very much the exception. This study set out to identify whether selection ratings and applicant characteristics predict performance during clinical psychology training. In considering the results, it should be borne in mind that in view of the high applicant to place ratio (average 28:1), the data presented here are very positively skewed as they only pertain to those successful in gaining a place; other than for A-level results, data variance was relatively small. It was not possible to make any corrections for range restriction. The actual predictive validity of the selection variables considered is probably higher than the observed relationships. The highly selective nature of the course makes it harder to see relationships between selection variables and outcomes, effectively reducing power. More research is needed to address this issue.

The key findings can be summed up as follows: generally, performance on one part of the course was correlated with performance on other parts of the course, with exam results showing statistically significant small to medium correlations with clinical placement concerns, viva outcome, and case report marks. However, against expectations, case reports correlated with academic, not clinical, performance, raising questions about the validity of case reports as indicators of clinical performance (cf. Simpson *et al.*, 2010). From all the information available at selection, school leaving exam grades (A-levels) were the most important predictor of performance during training; as noted, they were also the only data that showed a reasonable range. They predicted marks on all four of the exams independently of other pre-course variables, and were univariately associated with clinical placement problems. This corroborates evidence from medicine where A-levels have been found to predict academic performance many years after graduation (McManus, Smithers, Partridge, Keeling, & Fleming, 2003). While caution has been urged about the use of A-levels in selection, given that they are influenced by social and educational advantage (Scior *et al.*, 2007), in this study A-levels had a clear role in predicting performance. Although there was less variance in degree scores than in A-level scores, degree performance also predicted exam performance independently from A-levels on year 1 but not year 2 exams. University type was predictive of performance on the year 1 research methods exam and the year 2 statistics exam, with students who attended a post-1992 institution or a university outside the United Kingdom performing worse, and Oxbridge students performing better, on the statistics exam.

Demographic factors were also predictive: non-white students performed worse in the year 1 research methods exam and the year 2 statistics exam. Age also independently predicted the statistics exam. Trainees were relatively diverse in terms of age, but only a

small proportion were males (15%) and an even smaller proportion (9%) were from BME backgrounds. While the proportion of BME trainees compares fairly well with the 10% of people from BME backgrounds nationally (Office for National Statistics, 2011), as a London based course it compares poorly with the 34% of the Greater London population (Greater London Authority, 2011). The relationship observed in this sample between ethnicity and performance raises the concern that the underperformance of non-white students seen in medical education (Woolf *et al.*, 2011) and more broadly in higher education (Richardson, 2008) may also be seen in clinical psychology. This warrants further investigation, particularly as the UK Equality Act 2010 places a duty on all public authorities, including universities and the NHS, to monitor admission and progress of students by ethnic group to be able to address inequalities or disadvantage.

The finding that those with more clinical experience did worse in the statistics exam than those with minimal clinical experience is likely to be because the factors that impeded their exam performance also caused them to take longer to gain a training place, and thus they had more time to gain clinical experience.

Retrospective research supervisor ratings correlated with all outcomes and predicted case report marks and three of the four exam marks suggesting that these may measure global course ability, rather than just research skills. In contrast, contemporary interview ratings and referee ratings were not generally predictive of performance. The exception was with a marginal relationship with one of the year 1 exams. The results here were complicated: when we added the references to the regression model, both the academic and clinical reference were *negative* predictors; this may be a spurious relationship, however. None of the information available at selection predicted case report performance in our multivariate analyses.

The demographics of the trainee cohorts studied were very similar to the national picture of an average female: male ratio of 8.5:1 and a white: BME ratio of 9:1 (Clearing House data for 2002–2008). It is not possible to directly compare the academic qualifications of the present cohorts to the national picture as national data for the relevant period only records undergraduate results for those without post-graduate qualifications. However, given the high applicant: place ratio, those with first class degrees (40% across the cohorts studied) may be overrepresented (21.7% nationally had a first class degree, but many of the 25.4% of trainees nationally with Masters and PhD qualifications may have also had a first). This suggests that the findings are of relevance to other training providers in the United Kingdom. Overall, in view of our data, trainers and selectors should consider paying attention to A-levels and to some extent degree mark and university type in reaching decisions about who is likely to perform well on clinical psychology courses. This may seem a very unwelcome conclusion and raises ethical concerns. The desire to balance selecting students who are likely to perform well during training must be weighed carefully against the desire to select a diverse student body and profession. Furthermore, evidence that individuals from BME backgrounds are over-represented in less highly regarded universities (Shiner & Modood, 2002; Turpin & Fensom, 2004) suggests that increased attention to applicants' academic history may run counter to attempts to widen access to the profession. One message does emerge clearly though from the findings: while references and interviews clearly play a crucial role in selection, they do not appear to predict actual performance during clinical training. This may well be because they help deselect unsuitable applicants, thus reducing the trainee body to individuals likely to broadly perform well, as suggested by very low drop-out and failure rates. However, those involved in reaching selection decisions may well wish to reconsider what relative

importance they pay to the range of information available about applicants seen for interview.

Limitations

The study findings relate only to one course, albeit the largest in the United Kingdom. Given that selection processes and criteria vary across courses, findings may not generalize to other settings. Furthermore, this was an exploratory study and we performed a large number of analyses; the possibility of type 1 error should be borne in mind.

Our analyses relied on quantitative indicators that could be accessed. Many other variables, not least personality factors and life events, may well contribute to performance during training but were not measured here. Furthermore, many of the analyses relied on retrospective ratings, which may be unreliable. Research supervisors generally had fairly extensive contact with trainees under their supervision and made use of the full 5-point scale, indicating that they were able to recall trainees' performance fairly well. However, they suffer from the usual limitations of subjective ratings; due to the one-to-one supervisor–trainee relationship, it was not possible to assess inter-rater reliability. Course tutor ratings should be viewed with caution; due to the time lag involved, their reliability is questionable. It may be advisable for courses to collect contemporary ratings of trainee performance beyond academic and placement indicators, and to test their reliability and usefulness in monitoring trainee progress. The significant limitations of the data we had to rely on suggests that the reliability and validity of performance indicators commonly used during clinical psychology training merit careful consideration. Future research of this type should aim to use prospective data and more robust measures.

Conclusions

We want our selection methods to be as fair as possible a way of selecting among candidates who have all passed two previous stages of selection and who present relatively similar achievements to date. Our hope that interviews, and the applicants' references, would predict performance were disappointed. We presume that the interview process screens out unsuitable candidates given that drop-out and failure are very much the exception. The range of scores of both references and interview ratings was small for accepted applicants, so it is unclear whether the finding that they did not predict performance is to do with inadequate variance in the data, poor predictive power of the judgements which give rise to the ratings (cf. Stanton & Stephens, 2012), range restriction, or perhaps because interviews elicit valuable information which is nevertheless not then summatively evaluated during training.

Dropout and failure rates were very low, with all but two trainees who dropped out early completing their training successfully, even where certain aspects of training had to be repeated due to initial failure. Only 4% failed a case report, 1% a placement, and none failed their thesis viva. Further research is needed to understand how performance on the course relates to practice over the longer term as a clinical psychologist, and whether all those who complete training are indeed fit to practice. In medicine, it is commonly asserted that being a 'good' doctor is not (just) about performance in exams, and that other factors that are harder to measure and changeable make the difference between good, poor and average doctors (Journal British Medical, 2002). The same is probably true of clinical psychologists. With so many strong students applying for so few places, it is worth asking whether any selection methods can choose which students will perform best as

clinical psychologists after they qualify. Would a lottery system choosing among all students judged to meet entry criteria be fairer to applicants, trainees, and ultimately to service users (cf. Simpson, 1975)? Or are current attempts to develop a national screening test a move in the right direction?

Acknowledgements

We thank Julia Curl for her generous support in compiling data for this study, and Pasco Fearon and Tony Roth for their comments on a manuscript draft.

References

- Burt, C. (1943). Validating tests for personnel selection. *British Journal of Psychology*, *34*, 1–19. doi:10.1111/j.2044-8295.1943.tb01065
- Cherry, D. K., Messenger, L. C., & Jacoby, A. M. (2000). An examination of training model outcomes in clinical psychology programs. *Professional Psychology: Research & Practice*, *31*, 562–568. doi:10.1037/0735-7028.31.5.562
- Donnon, T., Paolucci, E. O., & Violato, C. (2007). The predictive validity of the MCAT for medical school performance and medical board licensing examinations: A meta-analysis of the published research. *Academic Medicine*, *82*, 100–106. doi:10.1097/01.ACM.0000249878.25186.b7
- Elston, M. A. (2009). *Women and medicine: The future*. London, UK: Royal College of Physicians of London.
- Emery, J.L., Bell, J.F., & Vidal Rodeiro, C.L. (2011). The biomedical admissions test for medical student selection: Issues of fairness and bias. *Medical Teacher*, *33*, 62–71. doi:10.3109/0142159X.2010.528811
- Eva, K. W., Reiter, H. I., Trinh, K., Wasi, P., Rosenfeld, J., & Norman, G. R. (2009). Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education*, *43*, 767–775. doi:10.1111/j.1365-2923.2009.03407
- Eva, K. W., Rosenfeld, J., Reiter, H. I., & Norman, G. R. (2004). An admissions OSCE: The multiple mini-interview. *Medical Education*, *38*, 314–326. doi:10.1046/j.1365-2923.2004.01776
- Ferguson, E., James, D., & Madeley, L. (2002). Factors associated with success in medical school: Systematic review of the literature. *British Medical Journal*, *324*, 952–957. doi:10.1136/bmj.324.7343.952
- Ferguson, E., James, D., O'Hehir, F., Sanders, A., & McManus, I. C. (2003). Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degree. *British Medical Journal*, *326*, 429–432. doi:10.1136/bmj.326.7386.429
- Goho, J., & Blackman, A. (2006). The effectiveness of academic admission interviews: An exploratory meta-analysis. *Medical Teacher*, *28*, 335–340. doi:10.1080/01421590600603418
- Greater London Authority (2011). *GLA Intelligence Update: 2010 Round Ethnic Group Population Projections using the 2009 Strategic Housing Land Availability Assessment (SHLAA)*. London, UK: Author.
- Hemmings, R., & Simpson, J. (2008). *Investigating the predictive validity of the Lancaster DClinPsy written shortlisting test on subsequent trainee performance: Final Report to the Clearing House*. Doctorate in Clinical Psychology: Lancaster University.
- Higham, J., & Steer, P. J. (2004). Gender gap in undergraduate experience and performance in obstetrics and gynaecology: Analysis of clinical experience logs. *British Medical Journal*, *328*, 142–143. doi:10.1136/bmj.328.7432.142
- Journal British Medical (2002). What's a good doctor and how do you make one? *British Medical Journal*, *325*, 711. doi:10.1136/bmj.325.7366.711

- Mathers, J., Sitch, A., Marsh, J. L., & Parry, J. (2011). Widening access to medical education for under-represented socioeconomic groups: Population based cross sectional analysis of UK data, 2002-6. *British Medical Journal*, *342*, d918. doi:10.1136/bmj.d918
- McFall, R. M. (2006). Doctoral training in clinical psychology. *Annual Review Clinical Psychology*, *2*, 21–49. doi:10.1146/annurev.clinpsy.2.022305.095245
- McManus, I.C., Dewberry, C., Nicholson, S., Dowell, J.S., Woolf, K., & Potts, H.W.W. (in press). *Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: Meta-regression of six UK longitudinal studies*. BMC Medicine.
- McManus, I. C., Ferguson, E., Wakeford, R., Powis, D., & James, D. (2011). Predictive validity of the Biomedical Admissions Test: An evaluation and case study. *Medical Teacher*, *33*, 53–57. doi:10.3109/0142159X.2010.525267
- McManus, I. C., Powis, D. A., Wakeford, R., Ferguson, E., James, D., & Richards, P. (2005). Intellectual aptitude tests and A-levels for selecting UK school leaver entrants for medical school. *British Medical Journal*, *331*, 555–559. doi:10.1136/bmj.331.7516.555
- McManus, I. C., Smithers, E., Partridge, P., Keeling, A., & Fleming, P. R. (2003). A-levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *British Medical Journal*, *327*, 139–142. doi:10.1136/bmj.327.7407.139
- Mercer, A., & Puddey, I. B. (2011). Admission selection criteria as predictors of outcomes in an undergraduate medical course: A prospective study. *Medical Teacher*, *33*, 997–1004. doi:10.3109/0142159X.2011.577123
- Office for National Statistics (2011). *Population estimates by ethnic group*, London, UK: Author.
- O'Neill, L.D., Wallstedt, B., Eika, B., & Hartvigsen, J. (2011). Factors associated with dropout in medical education: A literature review. *Medical Education*, *45*, 440–454. doi:10.1111/j.1365-2923.2010.03898
- Parry, J., Mathers, J., Stevens, A., Parsons, A., Lilford, R., Spurgeon, P., & Thomas, H. (2006). Admissions processes for five year medical courses at English schools: Review. *British Medical Journal*, *332*, 1005. doi:10.1136/bmj.38768.590174.55
- Peskun, C., Detsky, A., & Shandling, M. (2007). Effectiveness of medical school admissions criteria in predicting residency ranking four years later. *Medical Education*, *41*, 57–64. doi: 10.1111/j.1365-2929.2006.02647
- Phillips, A., Hatton, C., & Gray, I. (2004). Factors predicting the short-listing and selection of trainee clinical psychologists: A prospective National Cohort Study. *Clinical Psychology & Psychotherapy*, *11*, 111–125. doi:10.1002/cpp.399
- Richardson, J. T. E. (2008). *Degree attainment, ethnicity and gender: A literature review*. London, UK: The Higher Education Academy.
- Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, *85*, 112–118. doi:10.1037/0021-9010.85.1.112
- Scior, K., Gray, J., Halsey, R., & Roth, A. (2007). Selection for clinical psychology training: Is there evidence of any bias towards applicants from ethnic minorities? *Clinical Psychology Forum*, *175*, 7–11.
- Shiner, M., & Modood, T. (2002). Help or hindrance? Higher education and the route to ethnic equality. *British Journal of Sociology of Education*, *23*, 209–231. doi:10.1080/01425690220137729
- Simpson, M. A. (1975). Letter: Selection of medical students by lottery. *Journal of Medical Education*, *50*, 417–418.
- Simpson, J., Hemmings, R., Daiches, R., & Amor, C. (2010). Shortlisting from the Clearing House Application Form: Is it fit for purpose? *Psychology Learning & Teaching*, *9*, 32–36. doi:10.2304/plat.2010.9.2.32
- Siu, E., & Reiter, H. I. (2009). Overview: What's worked and what hasn't as a guide towards predictive admissions tool development. *Advances in Health Sciences Education*, *14*, 759–775. doi:10.1007/s10459-009-9160-8
- Stanton, A., & Stephens, C. (2012). Should all prospective medical students be interviewed? No. *Student British Medical Journal*, *20*, e3508. doi:10.1136/sbmj.f3950

- Turpin, G., & Fensom, P. (2004). *Widening access within undergraduate psychology education and its implications for professional psychology: Gender, disability and ethnic diversity*. Leicester, UK: British Psychological Society.
- Wilkinson, D., Zhuang, J., Byrne, G. J., Luke, H., Ozolins, I., Parker, M. H., & Peterson, R. F. (2008). Medical school selection criteria and the prediction of academic performance. *Medical Journal of Australia*, *188*, 349–354.
- Woolf, K., McManus, I. C., Potts, H. W., & Dacre, J. (2013). The mediators of minority ethnic underperformance in final medical school examinations: A longitudinal study. *British Journal of Educational Psychology*, *83*, 135–159. doi:10.1111/j.2044-8279.2011.02060
- Woolf, K., Potts, H., & McManus, I. C. (2011). The relationship between ethnicity and academic performance in UK-trained doctors and medical students: A systematic review and meta-analysis. *British Medical Journal*, *342*, d901. doi:10.1136/bmj.d901
- Yates, J., & James, D. (2006). Predicting the “strugglers”: A case-control study of students at Nottingham University Medical School. *British Medical Journal*, *332*, 1009–1013. doi:10.1136/bmj.38730.678310.63
- Yates, J., & James, D. (2010). The value of the UK Clinical Aptitude Test in predicting pre-clinical performance: A prospective cohort study at Nottingham Medical School. *BMC Medical Education*, *10*, 55. doi:10.1186/1472-6920-10-55

Received 8 October 2012; revised version received 17 September 2013