Received: 18 February 2020 | Revised: 15 September 2020 | Accepted: 21 September 2020

DOI: 10.1111/ijsa.12312

FEATURE ARTICLE

International Journal of Selection and Assessment | WILEY

Should research experience be used for selection into graduate school: A discussion and meta-analytic synthesis of the available evidence

Anthony Miller | **Marcus Crede** | **Lukas K. Sotola**

Department of Psychology, Iowa State University, Ames, IA, USA

Correspondence
Anthony Miller, Iowa State University, W112 Lagomarcino Hall, Ames, IA 50011, USA.
Email: milleraj@iastate.edu

Abstract

Prior research experience is widely considered by graduate school admissions committees in the United States of America. Here, we use meta-analytic methods and data from 18 unique samples and a total sample size of 3,525 students to shed light on the validity of prior research experience as a predictor of graduate school performance. Prior research experience was largely unrelated to academic performance (ρ = .01, k = 8, N = 1,419), degree attainment (ρ = .05, k = 3, N = 140), professional/practice performance (ρ = .06, k = 4, N = 1,120), and publication performance (ρ = .11, k = 7, N = 1,094). We also discuss whether consideration of prior research experience may unfairly disadvantage the students with lower levels of SES, students with childcare or eldercare responsibilities, and students from institutions at which research opportunities are limited.

KEYWORDS

fairness/bias/adverse impact, selection/placement

1 | INTRODUCTION

Selecting students into graduate school is a high-stakes decision. Universities invest large amounts of money into PhD students in the form of stipends, tuition waivers, lab equipment, and instructional costs. Indeed, the University of California (2018) estimated that graduate students in the Health Sciences cost an average of over $200,000 per student per annum, with somewhat lower but still considerable costs estimated for other disciplines. Graduate students, of course, also make substantial investments of their own in graduate school, in the form of time and the foregone wages that could have been earned in other pursuits. Admitting students who are most likely to succeed in graduate school is, therefore, in the interests of students, universities, and the taxpayers and funding agencies who fund many of these institutions.

To maximize the accuracy of admissions decisions, universities rely on a variety of materials. These vary widely across countries, institutions, and programs, but in the United States of America typically include admissions test scores (e.g., GRE scores), undergraduate grades, letters of recommendation, personal statements, and prior research experience. Meta-analytic syntheses have shown that undergraduate grades and scores on widely used admissions tests such as the GRE are relatively strongly related to the academic performance in graduate school and moderately related to other indicators of graduate student performance such as publication record, performance on licensing exams, and degree completion (see Kuncel & Hezlett, 2007 for a summary of this literature).

In recent years an increasing number of graduate programs have stopped relying on the GRE for making admissions decisions. This decision appears to be based on concerns about the cost of the GRE and possible test bias as well as the belief that the GRE is largely unrelated to performance in graduate school (Langin, 2020; Urry, 2015). We disagree with this assessment both because of clear meta-analytic evidence that GRE scores are predictive of graduate school performance, and because this new skepticism appears to be based on local validity studies that have failed to take into account the substantial range restriction of GRE scores in samples of admitted graduate students (e.g., Moneta-Koehler et al., 2017) or that have inappropriately aggregated data across institutions of different levels of selectivity (e.g., Miller et al., 2019).

That being said, this newly emergent skepticism about the validity of the GRE makes it even more important to clarify the validity of other widely used predictors of graduate student performance. Undergraduate grades have been shown to be excellent predictors (see Kuncel et al., 2001), while letters of recommendation are less predictive of academic performance but may hold some promise if interpreted appropriately (see Kuncel et al., 2014). The value of one other predictor of graduate school performance––prior research experience––has not been well established despite the fact that it is one of the most widely considered graduate school admissions criteria. For example, research experience was rated as the most important criterion after undergraduate grades, GRE scores, and letters of recommendation by 180 psychology faculty (Keith-Spiegel et al., 1994) and also 55 admissions committee chairs (Landrum et al., 1994), while a later survey of graduate programs in psychology (Landrum & Clark, 2005) found that research experience was, on average, rated as more important for the admissions process than the GRE scores. In this paper, we describe a meta-analytic examination of the validity of prior research experience for predicting graduate school performance. Specifically we aim to provide meta-analytic evidence for the predictive validity of prior research experience for four indicators of graduate school performance: (a) academic performance, (b) performance on indicators of professional practice, (c) publication rates, and (d) degree attainment. In order to shed further light on the degree to which a consideration of prior research experience may have value in the graduate school admissions process, we also examine the relationship of prior research experience with other predictors of graduate school success, such as GRE scores and undergraduate GPA.

2 | LITERATURE REVIEW

We begin by considering the theoretical reasons why a consideration of prior research experience in the graduate school admissions process is popular and may appear to be reasonable, and then, discuss plausible reasons why a consideration of prior research experience may not be advisable.

Research experience may predict performance in graduate school for at least four broad reasons––the first two of which are reflected in Campbell's Model of Job Performance (Campbell et al., 1993). This model describes the three proximal determinants of performance: declarative knowledge, procedural knowledge, and motivation. Research experience may result in the acquisition of job-relevant declarative knowledge (e.g., knowledge about biochemistry for an aspiring biochemistry graduate student) and job-relevant procedural knowledge (e.g., knowledge about the procedures that are used in biochemistry research). If we view research experience in a research laboratory as a form of training, then we would expect this experience to result in higher levels of declarative knowledge and procedural knowledge. This benefit of undergraduate research experience was demonstrated by Craney et al. (2011) who found that undergraduate research experiences were associated with increased levels of communication skills, problem solving skills, and interest in advanced study, while Harsh et al. (2011) found that students reported the development of basic lab techniques as one of the biggest benefits of participating in research activities as an undergraduate (see also Chemers et al., 2011; Landrum & Clark, 2005). The higher levels of declarative and procedural knowledge that are associated with participation in undergraduate research activities should, in turn, result in better performance in graduate school.

Research experience may also act as an indicator of motivation. That is, applicants who have engaged in prior research could reasonably be inferred to have higher levels of motivation for graduate school and research than applicants who did not engage in prior research. Importantly, prior research is relatively difficult to fake because letters of recommendation are often written by faculty who supervised a student in undergraduate research and, therefore, also typically attest to the fact that an applicant did, in fact, take part in prior research activities. Moreover, the amount of prior research experience––as opposed to the mere existence of research experience––may reasonably be considered during the admissions process, because an applicant who has been part of a lab for 2 years presumably has acquired more knowledge and exhibits greater motivation than an applicant who has only worked in a lab for a single semester or summer.

Exposure to research may also provide students with a realistic preview of graduate school. That is, applicants are likely to be more aware of the activities that graduate students engage in, the skills required to excel as a graduate student, and whether or not they find the work activities enjoyable and rewarding. Realistic job previews have been linked to both higher performance and lower attrition in a wide variety of occupations (see Phillips, 1998 for a meta-analytic review) and it is, therefore, reasonable to assume that applicants who have had exposure to the research process would exhibit higher performance in graduate school and be less likely to drop out. Findings from Harsh et al. (2011) who found that exposure to the scientific research process was cited as the biggest benefit of working in research laboratories as an undergraduate student would suggest that undergraduate research experiences can provide realistic previews of the work that graduate students engage in.

Finally, extensive prior research experience may also indicate that an applicant is able to work effectively with others and that the applicant has been able to exhibit performance in a research setting that was high enough so that the applicant was able to continue in that position. That is, research experience for more than a single semester may indicate a minimum level of competence in research.

Moreover, there are also reasons why it may be inadvisable for admissions committees to consider candidates' prior research experience. First, the tasks that undergraduate research assistants engage in—for example, administering experimental materials, coding simple data—may not be comparable to the research activities that they might engage in as graduate students—for example, designing studies, analyzing data, and writing manuscripts.

As such, the declarative knowledge and procedural knowledge that are acquired as an undergraduate research assistant may have only limited relevance for the job of a graduate student. Second, at many institutions, undergraduates can work as research assistants for course credit, and it may, in fact, be easier to earn credits while working in a lab than by taking demanding upper-level courses. Thus, some unmotivated students may decide to take research credits to avoid taking more challenging classes, which suggests that research experience may actually be *negatively* related to motivation. Third, it is difficult to place prior research experiences on a common-metric. The process of selecting from a pool of applicants requires that applicants are compared to each other. However, research experiences are likely to vary substantially from student to student, from supervisor to supervisor, from subject area to subject area, and from institution to institution, making comparison between applicants difficult. Fourth, faculty may select undergraduate research assistants on the basis of their grades (e.g., taking students who have the highest GPA from a pool of applicants). If this is the case, then considering research experience may have little predictive power above and beyond undergraduate GPA. That is, research experience may merely be an unintentional secondary indicator of GPA rather than providing any unique information about the applicant. Fifth, there is some evidence that the gain in research skills may be very limited for research experiences that are relatively short, such as the research boot camps examined by Feldon et al. (2016). Finally, there is good meta-analytic evidence that prior work experience is a relatively poor predictor of both general job performance and turnover (Van Iddekinge et al., 2019).

Thus, a meta-analytic study that takes into account as much research as possible on the relationship between research experience and graduate school performance will be a useful addition to the literature. It may clarify which of the above theoretical reasons for and against considering research experience in graduate admissions are most plausible, and it will help the admissions committees to decide how strongly they should weigh research experience alongside other predictors while considering graduate school applicants. Finally, it will contribute to the wider conversation surrounding the usefulness of various predictors of graduate school performance.

3 | METHOD

Sources for inclusion in this review were identified via keyword and abstract searches of the PsycINFO, ERIC, and Dissertation Abstracts International Databases. Search terms that were used are: “research experience” OR “prior research” OR “biodata” OR “lab” experience”, OR “research exposure” paired with “graduate OR PhD OR performance OR GPA OR success OR grades OR publication OR degree”. These searches were conducted by both the first and second author. These searches identified 4,353 potential source articles. The second author screened these articles for potential inclusion by considering their titles and abstracts. This initial screen resulted in 87 articles that were downloaded for closer examination to determine if the inclusion criteria were met. These initial searches were further supplemented by detailed searches of Google Scholar, asking academics over social media to supply relevant data, and examinations of the literature reviews and reference lists of sources identified via database searches. The second and third author conducted these supplemental searches. Approximately 50 further articles were considered for inclusion using this approach.

3.1 | Inclusion criteria

Sources were included if they reported the correlation between undergraduate research experience and either performance in graduate school or traditional predictors of graduate student success (e.g., GRE scores or undergraduate GPA). Sources were also included if they reported data in a form from which correlations could be calculated. The authors of four recently published studies that did not report correlations but that collected data on both undergraduate research experience and graduate school performance were contacted to request the correlations. Three of these four authors provided this information.

3.2 | Exclusion criteria

Studies were excluded if they reported on a subset of students that were also discussed by another study. In such cases, the data from the larger sample were coded. No studies were excluded because of the country in which the data were collected or the year in which the data were collected. One study was excluded because the authors aggregated data on degree completion across multiple different programs––such aggregation across programs can result in spurious effects due to Simpson's paradox. One study was excluded for only considering the very highest and very lowest performing students––such an extreme group design tends to produce inflated effect size estimates. We also excluded one study in which almost all participants had accumulated prior research experience in a master's program rather than as undergraduates because our focus was on the value of undergraduate research experience.

In some admission settings and some of the examined studies, prior research experience is treated as a dichotomous variable (e.g., research experience = 1, no research experience = 0). For this study we conceptualize the research experience as a continuous variable reflecting the amount of time that students have spent conducting research or working in laboratories. This is in line with both some prior studies on the relationship between research experience and graduate school performance (e.g., Mehrabian, 1970; Merolla & Serpe, 2013), and with the argument that research experience results in the acquisition of relevant procedural knowledge. That is, more exposure to research should result in a greater acquisition of procedural knowledge.

TABLE 1 Summary of coding for included studies

Authors	Discipline	Predictor (Scale)	Broad criterion category	*N*	*r*
Boudreau et al. (1983)	Psychology	Undergraduate research/teaching experience (0–2)	Academic performance	53	−.110
Boudreau et al. (1983)	Psychology	Undergraduate research/teaching experience (0–2)	Degree attainment	55	.020
Boudreau et al. (1983)	Psychology	Undergraduate research/teaching experience (0–2)	GRE-advanced	53	−.240
Boudreau et al. (1983)	Psychology	Undergraduate research/teaching experience (0–2)	GRE-Q	54	.090
Boudreau et al. (1983)	Psychology	Undergraduate research/teaching experience (0–2)	GRE-V	54	.130
Boudreau et al. (1983)	Psychology	Undergraduate research/teaching experience (0–2)	Letter of recommendation	54	.080
Boudreau et al. (1983)	Psychology	Undergraduate research/teaching experience (0–2)	UGPA	53	−.130
Dong et al. (2012)	Medicine	Research experience (Yes = 1, No = 0)	Academic performance	943	.025
Dong et al. (2012)	Medicine	Research experience (Yes = 1, No = 0)	UGPA	943	.013
Dong et al. (2012)	Medicine	Research experience (Yes = 1, No = 0)	Professional/practice performance	943	−.012
Feldon et al. (2016)	Life sciences	Duration of undergraduate research experience	Publication performance	298	.013
Gilmore et al. (2015)	STEM programs	Duration of undergraduate research experience	Academic performance	43	.285
Green and Bauer (1995)	Physical and Life Sciences	Research experience (1–5)	Academic performance	184	−.020
Green (1991)	Psychology & Bus. Admin.	Research experience (5 items)	Publication performance	88	.190
Hall et al. (2017)	Biomedical sciences	Duration of undergraduate research experience	GRE-Q	276	−.204
Hall et al. (2017)	Biomedical sciences	Duration of undergraduate research experience	GRE-V	276	−.065
Hall et al. (2017)	Biomedical sciences	Duration of undergraduate research experience	Letter of recommendation	247	−.018
Hall et al. (2017)	Biomedical sciences	Duration of undergraduate research experience	Publication performance	277	−.099
Hall et al. (2017)	Biomedical sciences	Duration of undergraduate research experience	UGPA	245	−.275
Izaak (2003)	Psychology	Undergraduate research experience (1 = Yes, 0 = No)	GRE-analytical	92	−.013
Izaak (2003)	Psychology	Undergraduate research experience (1 = Yes, 0 = No)	GRE-Q	92	−.116
Izaak (2003)	Psychology	Undergraduate Research Experience (1 = Yes, 0 = No)	GRE-V	92	−.218
Izaak (2003)	Psychology	Undergraduate research experience (1 = Yes, 0 = No)	UGPA	92	.295
Izaak (2003)	Psychology	Research experience (1–3)	Professional/practice performance	92	−.200
Kanna et al. (2009)	Medicine	Prior research experience (1 = Yes, 0 = No)	Clinical performance	51	.000
Mehrabian (1970)	Psychology	Duration of Undergraduate Research Experience	GRE-advanced	350	.130
Mehrabian (1970)	Psychology	Duration of undergraduate research experience	GRE-Q	350	.010

(Continues)

TABLE 1 (Continued)

Authors	Discipline	Predictor (Scale)	Broad criterion category	N	r
Mehrabian (1970)	Psychology	Duration of undergraduate research experience	GRE-V	350	.000
Mehrabian (1970)	Psychology	Duration of undergraduate research experience	UGPA	350	.030
Merolla and Serpe (2013)	STEM	Duration of undergraduate research experience	UGPA	694	.163
Pacheco et al. (2015)	Biomedical sciences	Research experience (1–3)	Degree attainment	42	.370
Pacheco et al. (2015)	Biomedical sciences	Research experience (1–3)	GRE-analytical	56	.137
Pacheco et al. (2015)	Biomedical sciences	Research experience (1–3)	GRE-Q	56	.000
Pacheco et al. (2015)	Biomedical sciences	Research experience (1–3)	GRE-V	56	.071
Pacheco et al. (2015)	Biomedical sciences	Research experience (1–3)	UGPA	55	.167
Paglis et al. (2006)	Physical and life sciences	Prior research experience	GRE-Q	233	−.120
Paglis et al. (2006)	Physical and life sciences	Prior research experience	GRE-V	233	−.070
Paglis et al. (2006)	Physical and life sciences	Prior research experience	Publication performance	233	.365
Park et al. (2018)	Biomedical sciences	Research experience (1–3)	GRE-Q	72	−.115
Park et al. (2018)	Biomedical sciences	Research experience (1–3)	GRE-V	72	−.459
Park et al. (2018)	Biomedical sciences	Research experience (1–3)	Publication performance	72	.231
Park et al. (2018)	Biomedical sciences	Research experience (1–3)	Academic performance	58	−.105
Park et al. (2018)	Biomedical sciences	Research experience (1–3)	UGPA	72	−.317
Piercy et al. (1995)	Family therapy	Faculty ratings of research experience	Academic performance	34	.110
Piercy et al. (1995)	Family therapy	Faculty ratings of research experience	Professional/practice performance	34	−.550
Piercy et al. (1995)	Family therapy	Faculty ratings of research experience	Publication performance	34	.090
Sheehy	Physical and life sciences	Undergraduate research experience (1 = Yes, 0 = No)	Publication performance	92	.259
Wiggins et al. (1969)—Sample 1	Psychology	Undergraduate research experience (1 = Yes, 0 = No)	Academic performance	46	−.165
Wiggins et al. (1969)—Sample 2	Psychology	Undergraduate research experience (1 = Yes, 0 = No)	Academic performance	58	−.013
Wilkerson (2007)	Physics	Presentation or publication as undergraduate student (1 = Yes, 0 = No)	Degree attainment	43	−.231

Note: r = zero-order correlation between research experience variable and criterion variable. When predictor variables were dichotomously coded correlations were first disattenuated for artificial dichotomization.

3.3 | Criterion categories

Graduate school performance was organized into three broad categories: (a) academic performance which comprised data on GPA, performance in individual classes, degree attainment, and faculty ratings of academic performance, (b) performance on indicators of professional competence (e.g., performance ratings during a clinical internship), and (c) publication performance (i.e., number of publications).

3.4 | Correlate categories

Information was available on the relationship between research experience and six other variables that are also used in the admissions process. The six variables that were coded were: (a) undergraduate GPA, (b) GRE-Verbal scores, (c) GRE-Quantitative Scores, (d) GRE-Analytic scores, (e) GRE-Analytical scores, and (f) Letters of recommendation.

TABLE 2 Meta-analytic estimates of the relationship between prior research experience and the 10 examined criteria and correlates

Criterion/correlate	k	N	$\bar{r}$	SD_r	$\bar{\rho}$	SD_ρ	95% CI	80% CR
Academic performance	8	1 419	.01	.08	.01	.00	[−0.06, 0.08]	[0.01, 0.01]
Degree attainment (1 = Yes, 0 = No)	3	140	.05	.29	.05	.25	[−0.68, 0.77]	[−0.43, 0.52]
Professional/practice performance	4	1120	.04	.15	.06	.19	[−0.27, 0.39]	[−0.26, 0.37]
Publication Performance	7	1094	.11	.19	.11	.17	[−0.06, 0.29]	[−0.14, 0.36]
UGPA	8	2504	.03	.16	.03	.16	[−0.11, 0.18]	[−0.19, 0.26]
GRE-Q	7	1133	−.08	.10	−.09	.06	[−0.19, 0.01]	[−0.18, 0.00]
GRE-V	7	1133	−.07	.14	−.07	.12	[−0.20, 0.06]	[−0.24, 0.10]
GRE-analytical	2	148	.04	.10	.05	.00	[−0.94, 1.03]	[0.05, 0.05]
GRE-subject	2	403	.08	.18	.08	.17	[−1.57, 1.73]	[−0.44, 0.60]
Letters of recommendation	2	301	−.00	.05	−.00	.00	[−0.48, 0.48]	[−0.00, −0.00]

Note: CI = confidence interval around $\bar{\rho}$; CR = credibility interval around $\bar{\rho}$; k = number of studies contributing to meta-analysis; N = total sample size; $\bar{r}$ = mean observed correlation; SD_r = observed standard deviation of r; SD_ρ = standard deviation of ρ after removing variance due to sampling error and variability in reliability; $\bar{\rho}$ = mean true-score correlation. Correlations corrected using artifact distributions.

3.5 | Coding process

Coding and the extraction of effect sizes for source articles and dissertations were systematized using a simple coding schema. All articles and dissertations were coded by all three authors. The level of agreement was near perfect and the few minor discrepancies (e.g., sample size) were resolved via discussion. For two source articles there were disagreements about the sample sizes that should be used due to discrepancies in how the sample size was reported in different sections of the articles. There was also one discrepancy in the coding of a predictor variable reliability estimate. The coding of all examined variables for all studies included in this review is provided in Table 1.

A number of specific coding decisions were made that are worth highlighting. First, Hall et al. (2017) reported on data for two types of publication performance: first author publications and total number of publications. We coded only the correlation involving total number of publications because this was most similar to the other sources that reported on publication performance. Second, whenever sources reported on correlations involving facets of a criterion variable, we computed a unit-weighted composite of correlations if the correlations among the facets was provided. Third, two studies (Green & Bauer, 1995; Paglis et al., 2006) reported on the same data set and for these two studies we coded the correlation based on the larger sample size when the correlation was based on the same two variables. Finally, we included one study reported on the correlation between graduate school performance and a variable that reflected prior research and teaching experience, because we assumed that prior teaching experience would likely be so rare that the correlation primarily reflected prior research experience. The final database was comprised of 50 correlations based on 18 unique samples and a total sample size of 3,525 students. The database for the focal research questions regarding the relationship of research experience with graduate school performance was comprised of 22 correlations from 15 unique samples and a total sample size of 2,510 students.

3.6 | Analytic approach

We computed meta-analytic estimates of the relationship between research performance and both the three types of graduate school performance and the four admissions variables using the random-effects model approach described by Hunter and Schmidt (2004). The Psychmeta package in R (Dahlke & Wiernik, 2018) was used. Although research experience is likely to contain some measurement error because it is sometimes based on self-reports, we did not correct for unreliability in the research experience variable because of the lack of available reliability information, because admissions decisions are made with such imperfectly reliable information, and because some information on prior research experience is obtained from theoretically perfectly reliable transcripts (e.g., grade obtained while working in a research laboratory). For similar reasons we also did not correct for unreliability in the "degree attainment", and "publication performance" criteria. In order to facilitate a comparison with meta-analyses of other predictors of graduate school performance we did correct for unreliability in the "graduate school academic performance", the "professional/practice performance" criteria, the GRE scores, and undergraduate GPA. All of these values were derived by taking the mean reported reliabilities from the Kuncel et al. (2001) meta-analysis of the GRE: for graduate school academic performance and undergraduate GPA we used alpha = .83, for professional/practice performance we used alpha = .55, for GRE-Q we used alpha = .90, for GRE-V we used alpha = .92, for GRE-Analytical we used alpha = .90, and for GRE-subject we used alpha = .94. No corrections for unreliability were done for letters of recommendation due to a lack of available evidence. Research does suggest the reliability of letters of recommendation is at around .4 (Kuncel et al., 2014), however the observed effect in this case (r = .00) would not be affected by reliability corrections.

Data on research experience were artificially dichotomized for six samples, while data on graduate student performance were artificially dichotomized for two samples. Correlations involving these studies were first corrected for artificial dichotomization prior to the computation of meta-analytic estimates using the formula provided by Hunter and Schmidt (2004). This formula relies on the sample specific split in the sample scores that authors used to dichotomize the variable. Thus, scores that were split at the median (e.g., high vs. low) will result in a smaller attenuation of the correlation than a more extreme split (e.g., top 10% vs. bottom 90%). The specific formula for the attenuating factor "*a*" is given as:

$$\frac{\varphi(c)}{\sqrt{P(1-P)}}$$

Where P is the proportion in the high split, c is the point on the normal curve that divides the distribution into proportions P and (1-P) and $\varphi(c)$ is the ordinate of c. For example, for a median split, p = .50, c = 0, and $\varphi(c)$ = .40 such that "*a*" is .80. That is, a correlation is, on average, reduced by 20% if the scores on one of the variables are artificially dichotomized at the median. An estimate of what the correlation would have been in the absence of such artificial dichotomization would, therefore, involve dividing the observed correlation by "*a*". For one study the proportional split in research experience was not reported and we, therefore, used the sample size weighted average split reported in the other six studies (63-37) for this one sample.

In our reporting of meta-analytic estimates we focus on seven quantities: k is the number of independent samples on which the estimate is based; N is the total sample size across the k samples; r_{obs} is the sample size weighted mean observed correlation; SD_{obs} is the sample size weighted observed standard deviation in effect sizes; SD_{res} is the estimate of the standard deviation of effect sizes after removing the variability that can be attributed to sampling error; 10%CR and 90%CR are the lower and upper limits of the 80% credibility interval; and 2.5%CI and 97.5%CI are the lower and upper limits of the 95% confidence interval.

Meta-analytic researchers are often concerned about the presence of publication bias in the literature that is being reviewed because such publication bias (e.g., effects that are large or statistically significant are more likely to be published) is likely to result in upwardly biased effect size estimates. Publication bias can result in funnel plots that are asymmetric and various statistical tests of funnel plot asymmetry have been developed (e.g., Pustejovsky & Rodgers, 2019). For this manuscript we did not test for funnel plot asymmetry for two reasons. First, the relationship between research experience and graduate school performance was not the focus of any of the journal articles or dissertations that were included in the review. For example, Feldon et al. (2016) examined the effectiveness of a research boot camp on the acquisition of research skills and the relationship of undergraduate research experience and graduate school performance was not even directly reported in the paper. As such publication bias is unlikely to have resulted in a biased distribution of effect sizes. A cursory inspection of the reported effect sizes (Table 1) also shows that many of the reported effect sizes were very low. Second, the number of studies available for review was so small that the power of formal tests of funnel plot asymmetry would have been unacceptably low.

4 | RESULTS

Meta-analytic estimate for the examined relationships are presented in Table 2. Prior research experience was largely unrelated to academic performance (ρ = .01, k = 7, N = 1,419), degree attainment (ρ = .05, k = 3, N = 140), professional/practice performance (ρ = .06, k = 3, N = 1,069), and publication performance (ρ = .11, k = 7, N = 1,094). It is noteworthy that with the exception of the academic performance criterion the credibility intervals for the three other criteria were very wide suggesting that there may be situations in which research experience is a relatively useful predictor. Unfortunately, there was insufficient data to allow an exploration of what these situational moderators might be.

Research experience was also largely unrelated to any of the other commonly examined predictors of graduate school performance, including undergraduate GPA (r = .03, k = 8, N = 2,504), GRE-Quantitative scores (r = −.08, k = 7, N = 1,133) GRE-Verbal scores (r = −.07, k = 7, N = 1,133), GRE-Advanced scores (r = .08, k = 2, N = 403), GRE-Analytic scores (r = .04, k = 2, N = 148), and Letters of Recommendation (r = .00, k = 2, N = 301). It is important to note that the correlation between variables that are used to select individuals into a sample (e.g., different admissions variables) can be distorted by collider bias (e.g., Borgen, 2019). Collider bias affects the correlation between two variables when individuals are selected into the sample on the basis of both variables––as is common in selection settings. Thus, a correlation between two variables such as research experience and GRE scores may be quite different––and even of a different sign––in a sample of graduate students and in a sample of graduate school applicants. As such the correlations with these other predictors of graduate school performance should be interpreted with caution.

5 | DISCUSSION

Research experience is widely used to select graduate students from pools of applicants. This practice may hold much intuitive appeal for faculty who are invested in admitting students who are enthusiastic about research, highly motivated, and who have had some prior exposure to the realities of conducting research. However, the relatively limited available data suggests that research experience is largely unrelated to the performance of graduate students and that a consideration of research experience in the admissions process does not result in a meaningful improvement in the quality of admitted graduate students. It is our view that the burden of proof for a selection practice lies with its proponents, and our data suggests that this burden has not been met.

Indeed, it can easily be shown that the inclusion of an invalid predictor such as research experience can substantially reduce the predictive validity of an admissions system. Consider the simple example of an admissions committee that only considers undergraduate grades in the admissions process. The meta-analytic results reported by Kuncel et al. (2001) indicate that undergraduate grades correlate with graduate school grades at $\rho = .30$. Adding prior research experience and giving it equal weight as undergraduate grades (as some surveys of admissions committees suggests is commonly done) would lower the correlation between the admissions systems score and graduate school grades to approximately $\rho = .22$. A consideration of Taylor-Russell tables would suggest that this decline in the ability to predict performance can result in a substantial decline in the number of students who succeed in graduate school. For example, assume that 30% of all applicants to a graduate school would be "good" students if they were admitted. Furthermore, assume that the graduate school can only admit 10% of all of the students who apply. If the graduate school were just to use undergraduate grades in making their admissions decisions (with a criterion validity of $r = .30$) the school would end up with 50% of students who would be classified as "good". Giving equal weight to research experience would reduce this proportion to about 44%. This difference is not trivial when considering the tremendous resources that are devoted to the training of PhD students in most disciplines.

In our introduction we discussed five theoretical reasons why admissions committees should be cautious about given a large weight to the prior research experience of graduate school applicants. These reasons referred to both the potential poor criterion validity of the data as well as potential adverse impact. Our results cannot directly speak to which, if any, of these reasons are correct but we suspect that the general lack of criterion validity of research experience is the result of two broad factors. First, it is difficult to compare different types of research experience to each other. One student may have just spent a summer helping to do a literature review on a topic while another student may have spent years actively working in a lab, designing experiments, analyzing data, and helping to write manuscripts. The latter experience is far more likely to be useful as these students transition into graduate school, but our reading of the literature suggest that prior research experience is often artificially dichotomized (as was done in almost half the studies included in this review), suggesting that these two experiences would be given equal weight in the admissions process. Future research on the value of prior research experience in the admissions process should avoid such a dichotomization and also attempt to gather far more specific information on the type of research activities that students were exposed to, how well they performed these activities, and the length of time that they spent on these activities. That is, the quality of research experiences may be more predictive than the simple existence of such research experiences.

Second, research experience opportunities vary dramatically across institutions. Students at small liberal arts colleges may have far fewer opportunities to experience formal training in a research laboratory because faculty are more focused on teaching than research. These same students may have developed other, perhaps even more relevant, skills such as writing and critical reasoning that easily compensate for their lack of research experience.

Emphasizing prior research experience in the admissions process may also have an adverse impact on students from previously underrepresented groups. If work in research labs is unpaid, as appears to be relatively common in psychology, then the students who have to engage in paid work to support themselves may be more likely to be excluded. Such students, of course, are much more likely to come from socioeconomic backgrounds that are systemically underrepresented in graduate school and academia (American Council on Education, 2006; Carnevale & Strohl, 2011). Not all students are equally able to devote unpaid labor toward gaining research experience. Recent data from the U.S. Census Bureau indicates that 43% of full-time students and 81% of part-time students are employed in paid positions while attending college, with almost half of all part-time students working at least 35 hr per week (The Condition of Education, 2019). Furthermore, more than one in five college students are parents (Radwin et al. 2018). These data suggest that many students have to work in paid positions in order to be able to attend university and support themselves and their families, while some students have childcare or eldercare responsibilities that limit the amount of time that they can devote to unpaid research activities or restrict their availability to such a degree that they are not selected for such positions by lab supervisors (Bangera & Brownell, 2014). Research opportunities for undergraduate students are also not readily available at many colleges and universities, particularly at those where faculty devote the majority of their time to teaching or where resources are insufficient to support faculty research activities.

Giving a high weight to prior research experiences may, therefore, primarily benefit students from relatively wealthy families who attend institutions at which research opportunities are readily available. This is particularly relevant when considering that a number of authors have found that undergraduate research experience is given a relatively large weight in the admissions process. Keith-Spiegel et al. (1994) found that research experience is the most widely used criterion used to distinguish among strong applicants to clinical and counseling psychology programs (i.e., those with high undergraduate GPAs, good GRE scores, and excellent letters of recommendation), while Landrum and Nelson (2002) demonstrated that faculty believe that undergraduate research experiences prepare students for graduate school. Indeed, undergraduate research experience was ranked only after undergraduate GPA, admissions test scores, and letters of recommendation as an admissions criterion in a survey of 55 graduate school admissions committees in psychology departments (Landrum et al., 1994). A later survey of 186 graduate programs in psychology (Norcross et al., 2005) found that research experience was weighted even more heavily than GRE scores and only slightly less than undergraduate grades and letters of recommendation. Furthermore, any other biases that exist in how undergraduate students are selected into undergraduate research experiences may also be reproduced at the graduate school level when prior research experience is emphasized in the graduate school admissions process. That is, if research supervisors are biased against certain groups when selecting students to work in their research laboratory then this bias will be reified at the graduate school selection stage if prior research experience is heavily weighted.

Because research experience appears to be used in the admissions process the pool of admitted students is likely to have a narrower range of scores on this variable than the pool of graduate school applicants. This would, in turn, result in observed correlations between research experience and criteria that are attenuated. One limitation of our review is that the lack of available information prevented us from making corrections for this range restriction in the research experience variable. Readers should, therefore, consider that our estimates are likely to be downwardly biased, although we believe that this attenuation effect is likely to be modest. Six studies reported the exact proportion of admitted students that had no prior research experience, and across these six studies we computed (using a sample-size weight) that, on average, 37% of students had no prior research experience at all. That is, in these programs many students are being admitted without any prior research experience. It is also important to remember that the estimated impact of range restriction is strongly influenced by the size of the observed correlation––the effect being much larger for larger correlations (Bobko, 1983). To provide readers with some idea of how much bigger the observed estimates might be under different range restriction scenarios we computed the attenuation factor "a" (see Hunter & Schmidt, 2004) for a scenario of severe range restriction; that is, where the ratio of the sample standard deviation and the population standard deviation in the research experience variable is 0.5 (the sample standard deviation is half as big as the population standard deviation). Further if the size of the observed correlation is $r = .10$ (representative of the stronger effects observed in our review), then the value of "a" is only $a = .98$. That is, the observed correlation is, on average, only 2% smaller than the population correlation. For a larger observed correlation, say, $r = .3$ and the same amount of range restriction the value of "a" would be $a = .52$ such that the observed correlation is almost half of the estimated population correlation.

We acknowledge that this literature is relatively sparse; further local validity studies are likely to be valuable in order to establish the added value of considering research experience for graduate admissions. More published and accessible research in the predictive value of undergraduate research experience is needed. However, if admissions practices are to be based on the best currently available evidence, then our findings strongly suggest that a consideration of prior research experience is not only unwarranted but may even decrease the validity of the overall selection process.

ORCID

Anthony Miller https://orcid.org/0000-0002-0869-0689

REFERENCES

American Council on Education. (2006, May). Working their way through college: Student employment and its impact on the college experience. Retrieved from https://www.acenet.edu/Documents/IssueBrief-2006-Working-their-way-through-College.pdf

Bangera, G., & Brownell, S. E. (2014). Course-based undergraduate research experiences can make scientific research more inclusive. *Cbe—life Sciences Education*, *13*(4), 602–606. https://doi.org/10.1187/cbe.14-06-0099

Bobko, P. (1983). An analysis of correlations corrected for attenuation and range restriction. *Journal of Applied Psychology*, *68*(4), 584–589. https://doi.org/10.1037/0021-9010.68.4.584

Borgen, N. T. (2019). Collider bias (aka sample selection bias) in observational studies: Why the effects of hyperandrogenism in elite women's sport are likely underestimated. *British Journal of Sports Medicine*, *54*, 750–752. https://doi.org/10.1136/bjsports-2018-100054

Boudreau, R. A., Killip, S. M., MacInnis, S. H., Milloy, D. G., & Rogers, T. B. (1983). An evaluation of Graduate Record Examinations as predictors of graduate success in a Canadian context. *Canadian Psychology/Psychologie Canadienne*, *24*(3), 191–199. https://doi.org/10.1037/h0080730

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. *Personnel Selection in Organizations*, *3570*, 35–70.

Carnevale, A. P., & Strohl, J. (2011). Our economically polarized college system: Separate and unequal. *The Chronicle of Higher Education*, *58*(6).1–64.

Chemers, M. M., Zurbriggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *Journal of Social Issues*, *67*(3), 469–491. https://doi.org/10.1111/j.1540-4560.2011.01710.x

Craney, C., McKay, T., Mazzeo, A., Morris, J., Prigodich, C., & Groot, R. (2011). Cross-discipline perceptions of the undergraduate research experience. *The Journal of Higher Education*, *82*(1), 92–113. https://doi.org/10.1353/jhe.2011.0000

Dahlke, J. A., & Wiernik, B. M. (2018). psychmeta: An R package for psychometric meta-analysis. *Applied Psychological Measurement*, *43*(5), 415–416. https://doi.org/10.1177/0146621618795933

Dong, T., Durning, S. J., Gilliland, W. R., Waechter, D. M., Cruess, D. F., DeZee, K. J., Calloway, M., & Artino, A. R. (2012). Exploring the relationship between self-reported research experience and performance in medical school and internship. *Military Medicine*, *177*(Suppl_9), 11–15. https://doi.org/10.7205/MILMED-D-12-00232

Feldon, D. F., Maher, M. A., Roksa, J., & Peugh, J. (2016). Cumulative advantage in the skill development of STEM graduate students: A mixed-methods study. *American Educational Research Journal*, *53*(1), 132–161. https://doi.org/10.3102/0002831215619942

Gilmore, J., Vieyra, M., Timmerman, B., Feldon, D., & Maher, M. (2015). The Relationship between Undergraduate Research Participation and Subsequent Research Performance of Early Career STEM Graduate Students. *The Journal of Higher Education*, *86*(6), 834–863. http://dx.doi.org/10.1080/00221546.2015.11777386

Green, S. G. (1991). Professional entry and the adviser relationship: Socialization, commitment, and productivity. *Group & Organization Studies*, *16*(4), 387–407. https://doi.org/10.1177/105960119101600404

Green, S. G., & Bauer, T. N. (1995). Supervisory mentoring by advisers: Relationships with doctoral student potential, productivity, and commitment. *Personnel Psychology*, *48*(3), 537–562. https://doi.org/10.1111/j.1744-6570.1995.tb01769.x

Hall, J. D., O'Connell, A. B., & Cook, J. G. (2017). Predictors of student productivity in biomedical graduate school applications. *PLoS ONE*, *12*(1), e0169121. https://doi.org/10.1371/journal.pone.0169121

Harsh, J. A., Maltese, A. V., & Tai, R. H. (2011). Undergraduate research experiences from a longitudinal perspective. (RESEARCH and TEACHING). *Journal of College Science Teaching*, *41*(1), 84–91.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis* (2nd ed.). Sage Publications. https://doi.org/10.4135/9781412985031

Izaak, A. R. (2003). The prediction of clinical competence in clinical psychology graduate students. Doctoral dissertation, Chicago School of Professional Psychology.

Kanna, B., Gu, Y., Akhuetie, J., & Dimitrov, V. (2009). Predicting performance using background characteristics of international medical graduates in an inner-city university-affiliated Internal Medicine residency training program. *BMC Medical Education*, *9*, 42. https://doi.org/10.1186/1472-6920-9-42

Keith-Spiegel, P., Tabachnick, B. G., & Spiegel, G. B. (1994). When demand exceeds supply: Second-order criteria used by graduate school selection committees. *Teaching of Psychology*, *21*(2), 79–81. https://doi.org/10.1207/s15328023top2102_3

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance.. *Psychological Bulletin*, *127*(1), 162–181. http://dx.doi.org/10.1037/0033-2909.127.1.162

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, *315*(5815), 1080–1081. https://doi.org/10.1111/ijsa.12060

Kuncel, N. R., Kochevar, R. J., & Ones, D. S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment*, *22*(1), 101–107. https://doi.org/10.1111/ijsa.12060

Landrum, R. E., & Clark, J. (2005). Graduate Admissions Criteria in Psychology: An Update. *Psychological Reports*, *97*(2), 481–484. http://dx.doi.org/10.2466/pr0.97.2.481-484

Landrum, R. E., Jeglum, E. B., & Cashin, J. R. (1994). The decision-making processes of graduate admissions committees in psychology. *Journal of Social Behavior and Personality*, *9*(2), 239–248.

Langin, K. (2020, June 24). A wave of graduate programs drops the GRE application requirement. Retrieved from https://www.sciencemag.org/careers/2019/05/wave-graduate-programs-drop-gre-application-requirement

Mehrabian, A. (1970). Ability factors of candidates for graduate school. *American Psychologist*, *25*(6), 560–563 https://doi.org/10.1037/h0037863

Merolla, D. M., & Serpe, R. T. (2013). STEM enrichment programs and graduate school matriculation: The role of science identity salience. *Social Psychology of Education*, *16*(4), 575–597. https://doi.org/10.1007/s11218-013-9233-7

Miller, C. W., Zwickl, B. M., Posselt, J. R., Silvestrini, R. T., & Hodapp, T. (2019). Typical physics Ph.D. admissions criteria limit access to underrepresented groups but fail to predict doctoral completion. *Science Advances*, *5*(1), eaat7550. http://dx.doi.org/10.1126/sciadv.aat7550

Moneta-Koehler, L., Brown, A. M., Petrie, K. A., Evans, B. J., & Chalkley, R. (2017). The limitations of the GRE in predicting success in biomedical graduate school. *PLoS ONE*, *12*(1), e0166742. https://doi.org/10.1371/journal.pone.0166742

Norcross, J. C., Kohout, J. L., & Wicherski, M. (2005). Graduate study in psychology: 1971–2004. *American Psychologist*, *60*, 959–975. https://doi.org/10.1037/0003-066X.60.9.959

Pacheco, W. I., Noel, R. J. Jr, Porter, J. T., & Appleyard, C. B. (2015). Beyond the GRE: Using a composite score to predict the success of Puerto Rican Students in a biomedical PhD program. *Cbe—Life Sciences Education*, *14*(2), ar13. https://doi.org/10.1187/cbe.14-11-0216

Paglis, L. L., Green, S. G., & Bauer, T. N. (2006). Does adviser mentoring add value? A longitudinal study of mentoring and doctoral student outcomes. *Research in Higher Education*, *47*(4), 451–476. https://doi.org/10.1007/s11162-005-9003-2

Park, H. Y., Berkowitz, O., Symes, K., & Dasgupta, S. (2018). The art and science of selecting graduate students in the biomedical sciences: Performance in doctoral study of the foundational sciences. *PLoS ONE*, *13*(4), e0193901. https://doi.org/10.1371/journal.pone.0193901

Phillips, J. M. (1998). Effects of realistic job previews on multiple organizational outcomes: A meta-analysis. *Academy of Management Journal*, *41*, 673–690. https://doi.org/10.5465/256964

Piercy, F. P., Dickey, M., Case, B., Sprenkle, D., Beer, J., Nelson, T., & McCollum, E. (1995). Admissions criteria as predictors of performance in a family therapy doctoral program. *The American Journal of Family Therapy*, *23*(3), 251–259. https://doi.org/10.1080/01926189508251355

Pustejovsky, J. E., & Rodgers, M. A. (2019). Testing for funnel plot asymmetry of standardized effect sizes. *Research Synthesis Methods*, *10*(1), 57–71. https://doi.org/10.1002/jrsm.1332

Radwin, D., Conzelmann, J. G., Nunnery, A., Lacy, T. A., Wu, J., Lew, S., Wine, J., & Siegel, P. (2018). 2015–16 National Postsecondary Student Aid Study (NPSAS:16): Student Financial Aid Estimates for 2015–16 (NCES 2018-466). U.S. Department of Education. Washington, DC: National Center for Education Statistics. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018466

Sheehy, B. N. (2019). Support Received from the Dissertation Advisor and the Graduate Student Success of Doctoral Students Majoring in the Sciences. *Graduate Theses and Dissertations*, 1–169. https://scholarcommons.usf.edu/etd/7937/

The Condition of Education. (2019). https://nces.ed.gov/programs/coe/pdf/coe_ssa.pdf

The University of California. (2018). Report on expenditures for undergraduate and graduate instruction. https://www.ucop.edu/operating-budget/budgets-and-reports/legislative-reports/expenditures_for_undergraduate_and_graduateinstruction-9-21-18.pdf

Urry, C. (2015, December 9). President's Column: Rethinking the Role of the GRE. Retrieved from https://aas.org/posts/news/2015/12/presidents-column-rethinking-role-gre

Van Iddekinge, C. H., Arnold, J. D., Frieder, R. E., & Roth, P. L. (2019). A meta-analysis of the criterion-related validity of prehire work experience. *Personnel Psychology*, *72*(4), 571–598. http://dx.doi.org/10.1111/peps.12335

Wiggins, N., Blackburn, M., & Richard Hackman, J. (1969). Prediction of first-year graduate success in psychology: Peer ratings. *The Journal of Educational Research*, *63*(2), 81–85. https://doi.org/10.1080/00220671.1969.10883941

Wilkerson, T. (2007). The relationship between admission credentials and the success of students admitted to a physics doctoral program. Electronic Theses and Dissertations, 2004-2019. 3411. https://stars.library.ucf.edu/etd/3411

How to cite this article: Miller, A, Crede, M, Sotola, LK. Should research experience be used for selection into graduate school: A discussion and meta-analytic synthesis of the available evidence. *Int J Select Assess*. 2020;00:1–10. https://doi.org/10.1111/ijsa.12312