

A Meta-Analysis on the Predictive Validity of Graduate Record Examination (GRE) General Test

Mohammad Hasan Razmi^{1*}, Masoud Khabir², Shouket Ahmad Tilwani³

Received: 16 April 2021

Accepted: 21 June 2021

Abstract

Since its inception in 1949, over 1,500 studies have investigated the validity of the GRE General Test to predict its performance criteria in higher education (Klieger, Bridgeman, Tannenbaum, & Cline, 2016). The present review paper sought to examine the predictive validity of the GRE General Test. Factors affecting the predictive validity (e.g., range restriction, compensatory selection, criterion unreliability, substantive and artifactual moderators, bias in testing, coaching effects, socioeconomic status (SES), gender, and a host of other intervening factors such as motivation, communication skills, etc.) have been discussed. A brief overview of GRE revised General Test format is also presented. After an account of the related review of the literature, a critical commentary on the predictive validity of the GRE General Test has been discussed with an emphasis on the role of criterion unreliability and SES factor effects.

Keywords: Compensatory selection; criterion unreliability; Graduate Record Examination (GRE); predictive validity; range restriction

1. Introduction

The Graduate Record Examination (GRE) General Test is a standardized test of verbal (GRE-V), quantitative (GRE-Q), and analytical reasoning (GRE-A) that was principally intended to assist applicants seeking admissions to the universities in the United States of America. The original testing system, as well as its subsequent updates (the most recent of which was made in August 2011), have been extensively administered and investigated since 1949. The results have been published in prestigious academic publications (Powers, 2004).

Students applying to (post)graduate programs at international institutions are increasingly taking the GRE (Liu, Klieger, Bochenek, Holtzman, & Xu, 2016). More than 90% of PhD programs and over 80% of master's programs in the United States mandate GRE scores (Norcross, Hanych, & Terranova, 1996). Scores on the GRE are sometimes utilized to improve

¹ English Department, Yazd University, Yazd, Iran (*Corresponding author: hasanrazmi2000@gmail.com)

² Kerman University of Medical Sciences, Kerman, Iran

³ Department of English, College of Science and Humanities, Prince Sattam Bin Abdulaziz University, Al-Kharj, 11942, Saudi Arabia

decision-making processes for scholarships and sources of funding available to academic members (Rock & Adler, 2014). Students with greater GRE scores are more likely to obtain financial assistance from the institution in the form of grants, teaching/research fellowships, and, in certain circumstances, direct financial aid as part of their enrollment. This financial assistance enables participants to acquire a good start in their academic careers, publish more articles, obtain better job opportunities, and, in many ways, outshine students who have lower levels of financial support and have to spend a significant amount of time to struggle with financial concerns (e.g., teaching), tuition, and living expenses (Lerdau, & Avery, 2007).

GRE is also claimed to be a test of general cognitive abilities that displays a fine-grained portrait of procedural and declarative knowledge, and GRE results represent individual differences in terms of motivational attitudes, problem solving, and social skills to pursue higher education (McCloy, Campbell, & Cudeck, 1994). While declarative knowledge refers to understanding what to accomplish, procedural knowledge refers to understanding how to get things done, and motivation is characterized with doing a task to the best of one's ability (Kuncel, Hezlett, & Ones, 2001). According to Schmidt and Hunter (1993), another key theory explaining the GRE's power to explain and predict performance outcomes is that the GRE's assessment of general cognitive skills has positive and strong associations with functioning skills required for successful graduate school performance. Graduate students with higher functional knowledge are likely to work more efficiently and successfully as they progress through a practical program, even beyond their academic careers.

The GRE-V, GRE-Q, and GRE-A, according to McCloy et al. (1994), assess personal skills or talents that could have decisive impacts on later graduate performance through procedural and declarative thinking skills. McCloy et al. argued that procedural knowledge essential to graduate school achievement would include activities such as reading and summarizing a paragraph (GRE-V).

All the above-mentioned accounts on the virtue of the GRE and its diagnostic ability prediction would definitely stimulate the academic members in charge into making the best use of GRE scores to make better decisions about the selection from the higher education applicant pool. Furthermore, the GRE seems to go one step beyond and predict success in functional ability in accordance with procedural knowledge. That would be highly promising if GRE scores could predict success not only during the academic life, but also during academic afterlife! To substantiate such a bold claim, empirical evidence is highly needed. A growing body of research has investigated whether GRE exerts an influence in this regard (Walsh, Arslan, & Finn, 2021). The results, while promising in some cases, have been conflicting at times, due to the complexity of the nature of prediction per se. The predictive validity odyssey is full of unpredictable adventures to be explored by avid researchers.

This study, therefore, chiefly seeks to summarize the challenges and previous findings related to predictive validity of the GRE General Test. The findings of the related investigations available in the literature are also presented. The main tenets of the predictive validity in accordance with GRE are also discussed.

2. Review of Literature

In this section, first the definition of the predictive validity will be presented. Then factors influencing predictive validity estimates will be discussed and a brief account of the GRE revised General Test format will be given. Finally, the review of related literature will be presented.

2.1. The predictive validity definition

In evaluation studies, predictive validity is referred to the degree to which a score on an examination or a measurement scale predicts achievement on a certain criterion factor (Cronbach & Meehl, 1955). According to Fulcher and Davidson (2007, p. 5), “[p]redictive validity is the term used when the test scores are used to predict some future criterion, such as academic success.” Predictive validity is a subcategory of criterion-oriented validation criteria which evaluates relationship between a particular test and a criterion to which we wish to make predictions. There are some challenges related to the estimation of predictive validity. The following section delineates the criteria to which GRE scores are supposed to predict as well as the factors influencing the predictive validity.

2.2. Factors influencing predictive validity estimates

According to Powers (2004), investigations on validity are particularly prone to inconsistencies due to the reliance on limited sample size, implausible criteria assignment, overgeneralization based on limited unreliable sources of information, disregard for the effects of compensatory selection, and lack of consideration for range restriction estimation effects in the predictive and criterion measures.

With regard to range restriction (when participants are chosen from a bigger group of candidates—for instance, based on test results), it was known more than one century ago that drawing samples from a population could possibly diminish the correlational associations among variables by condensing the magnitude of a measure (Pearson, 1903). This limitation is common in college admission processes. In most cases, this has the consequence of underestimating the genuine association between GRE test results and some other indicator of success in the original population, i.e., the candidate pool. This type of range restriction could exert a significant impact on validity coefficients (Linn & Dunbar, 1982). Predictive validity of the graduate record examination with and without range restraints has been of interest among the researchers in the field (e.g., Huitema & Stein, 1993; Oldfield & Hutchinson, 1997).

Due to the extreme impoverishment that arises from the inclusion of poor and inaccurate measures of success, it is frequently thought necessary to rectify the error of measurement in criterion variables. Current researches generally aim at correcting for unreliability in the criteria, as long as the adjusted statistics and the original raw data values are reported. (Powers, 2004). The criteria definition, assignment, limitation, and examination are discussed under the discussion section of the present study.

Another variable, criterion unreliability, can possibly reduce validity estimates. The criteria by which GRE scores are to predict success are of paramount significance. In order to examine whether the GRE scores predict success, Kuncel et al. (2001), in their meta-analysis,

probed eight distinctive factors: graduate grade point average (GGPA), comprehensive examination grades, first year GGPA, faculty ratings bestowed by departments, completion of PhD or master’s programs, number of publication citations, time to complete educational programs, and number of publications (Figure 1).

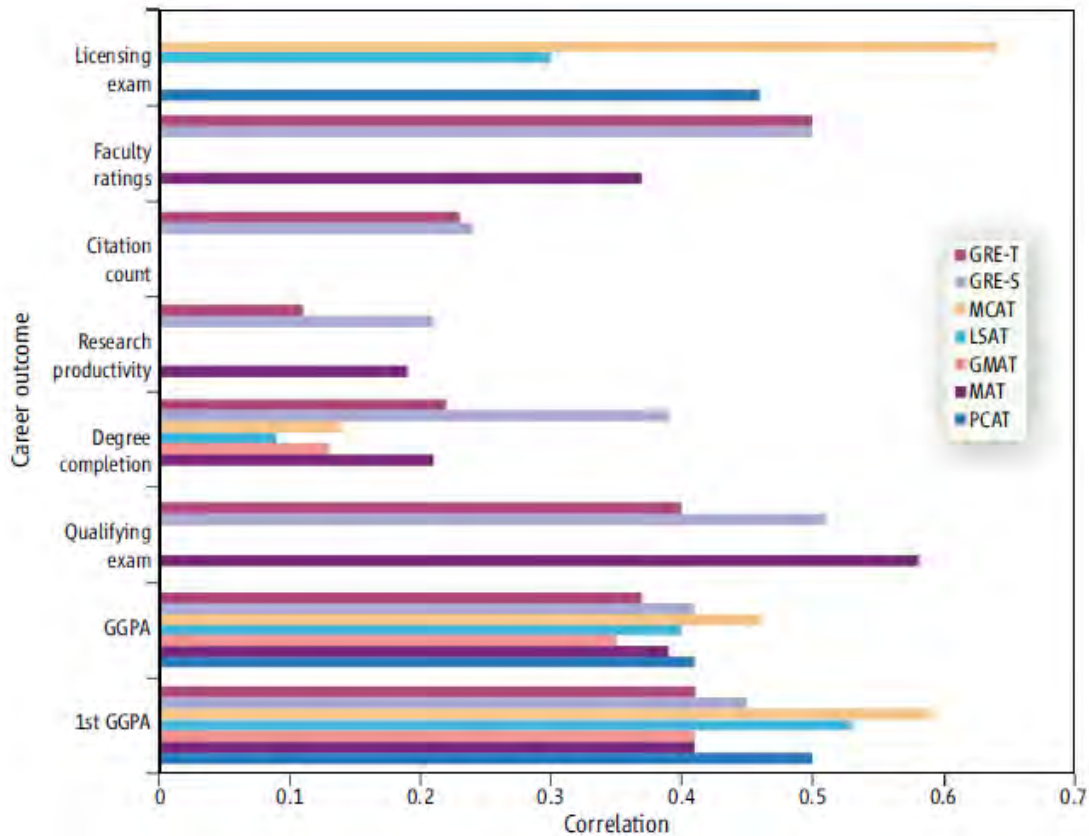


Figure 1. Academic success criteria in accordance with Standardized tests (Kuncel & Hezlett, 2007).

Another relevant factor is compensatory selection, which is defined as allowing successful performance on one selection criterion to make up for a poor score on another variable. This factor, similar to criterion unreliability and range restriction, can have a major effect on validity estimations.

Concerning the factors influencing the GRE predictive validity, Kuncel, Wee, Serafin, and Hezlett (2010) differentiate between two types of what they call moderators that may have impacts on the relationship between the GRE scores and an achievement outcome: substantive and artifactual. By substantive, they mean factors such as course complexity, under-developed educational settings, and discipline area. Artifactual type includes differential restriction of range and criterion measurement error differences.

Another threat is bias in assessment. A major consideration is that particular groups, including racial diversities, minorities, and gender groups, may be discriminated against in admission examinations (Wilkinson, Shugart, Williams, & Riechel, 2021). This calls for a careful examination of items of high-stakes tests in terms of differential item functioning (DIF) as well as any possible content bias. DIF specifically evaluates if the score on a single item of

a test varies across different ethnic, gender, and racial groups when total scores on a performance criterion of interest is controlled for (Amirian, Alavi, & Fidalgo, 2014). For example, Schwager, Hulsheger, Bridgeman, and Lang (2015) investigated GRE scores, socio-economic status, and college GPA as indicators of academic achievement. House (1994) examined how gender differences in GRE results could predict applicants' performance in college examinations.

Another idea pertinent to predictive validity is coaching effects in testing. Coaching refers to score improvement due to special preparation (Kuncel & Hezlett, 2007). Powers (1985) conducted a study on effects of test preparation on the validity of graduate admissions test. While acknowledging the detrimental effect of coaching, Powers, in his study, found no consistent effect of coaching on GRE scores.

According to Powers (2004, p. 209), other factors such as “motivation, maturity, dedication, citizenship, leadership, initiative, communication skills, research experience, animal contact, moral–ethical character, and a desire to serve the public” can be considered as criteria for success. These factors might come into play as intervening influences affecting the predictive validity of GRE scores. A high predictive validity coefficient must account at least partly for the above-mentioned dynamic variables within and across individuals.

2.3.A brief glance at the GRE revised General Test format

According to Educational Testing Service (2012), The revised format of the General Test of GRE was administered for the first time. This test is mainly used by colleges for student screening and applicant selection in the admission process. Some institutions tend to use the GRE scores to decide whether the applicants are eligible for financial support in the form of fellowship awards, research assistantship, teaching, etc. (Liu et al. 2016).

The administration of this standardized test generally takes around 3.5 to 4 hours. The test is composed of three main parts developed to evaluate a variety of cognitive abilities and skills: quantitative reasoning (GRE-Q), verbal reasoning (GRE-V), and analytical writing (GRE-A).

While GRE-Q and GRE-V are conducted in multiple-choice items, GRE-A utilizes constructed-response items. Both GRE-V and GRE-Q include two parts including 20 test items each. GRE-V examines the candidates' reasoning skills in comprehending what they read. GRE-Q, evaluates the candidates' skills in solving mathematical problems with an emphasis on data interpretation. The GRE revised General Test is a computer adaptive test (CAT) sensitive to the ability level of the examinees. It takes on a multi-step procedure in the second half of GRE-V and GRE-Q parts by incrementally assigning the candidates to one of three difficulty level packs of items (low, medium, and high) as they proceed through the examination according to the candidates' performance on previous sections of the exam. The application of this incremental multi-step procedure, according to Educational Testing Service (2012), can result in the accurate estimation of candidates' measured abilities.

GRE-A includes two constructed response writing tasks meant to evaluate candidates' abilities in (a) critically evaluating a problem and (b) convincingly supporting their ideas in the form of a set of written responses in English.

The GRE revised General Test is administered both in paper-based and computer-based versions, and nowadays, most of applicants prefer sit the computer-based format of the test. The computerized format allows the examinees to preview or (re)review items in a section and revise the given answers in each section. The GRE-Q in computer-based version, is equipped with an on-screen calculator to make calculations easier and to minimize the memory effect and mere attention to computational skills. Furthermore, the exam utilizes novel response types including numerical entries.

2.4. Review of the related literature

Over 1,500 researches have investigated the predictive validity of the GRE General Test in academic contexts (Klieger, Bridgeman, Tannenbaum, & Cline, 2016). Since its beginning in 1949, the GRE and its modified versions have been extensively administered and widely investigated (Powers, 2004). Following are the most prominent studies carried out by highly cited researchers related to the predictive validity of the examination.

Kuncel et al. (2001) launched a sizable meta-analysis including 1,753 participants. The results showed that both quantitative and verbal sections of GRE substantially predicted the candidates' GGPA, equivalent to the magnitude that UGPA could explain GGPA. The results also indicated that the GRE Subject Tests better predicted the GGPA compared to the GRE General test. The GRE Subject Tests were also strong predictors of participant' scores on the comprehensive exam. The quantitative section and GRE Subject Tests and the quantitative section of the General GRE moderately explained the number of publications cited.

Powers (2004) studied the predictive validity of the GRE scores in a sample of students admitted to the schools of veterinary. The required data were gathered from 16 colleges. This study included both admitted students and candidates to be admitted in college. Having corrected the range restriction estimates, and the criterion unreliability (the GPA obtained in first year of education), Powers found that GRE-Q and GRE-V had positive associations with students' first-year GPA scores. UGPA and GRE scores altogether explained almost 65% of the variability of students' GPA scores in their first year of education.

Burton and Wang (2005) explored the predictive validity of the verbal and quantitative sections of the GRE in terms of a number of criteria including range restriction, technical efficiency, expertise in field of study, total GPA scores, and communicative abilities across diverse fields of study including psychology, education, chemistry, biology, and English. The researchers conducted a number of correlational analyses to probe the scores obtained on the quantitative and verbal sections of the GRE in relation with the criterion variables of the study. The values of adjusted R corresponding to the criterion variables were between .17 to .66 across different academic discipline from psychology to education.

Kuncel et al. (2010) launched a meta-analysis on the predictive validity of GRE among ten thousand M.A. and PhD students participating in about 100 investigations. The criterion variables were ratings bestowed by departments, GGPA, and the GPA obtained in the first year of education at college. The findings indicated that that GRE scores were strong predictors of all three criterion variables in the study for both M.A. and PhD groups. However, GRE-V showed to be a better predictor of GGPA among the M.A. participants compared to their

counterparts in the PhD group. The researchers explained that the discrepancy might be due to the fact that the score range of doctoral students ($SD = .21$) was smaller than that of master students ($SD = .40$) in their study.

Klieger, Cline, Holtzman, Minsky, and Lorenz (2014) explored the effectiveness of the GRE scores in predicting GPA scores obtained by a large sample of master's and PhD students majoring in various disciplines at ten public colleges in Florida, USA. Besides its large sample (4,229 PhD and 21,127 M.A. students) including 28 subject areas, this research also utilized a variety of analytical methods. This investigation made a unique contribution to this area of research, since it was the first attempt to explore the predictive validity of the academic writing section of the GRE examination. The results indicated that the GRE-A scores significantly predicted both groups' GGPA scores (M.A. = .19 and PhD = .21). A point worthy of note is that this study controlled for the intervening effects of multivariate range restriction; however, the error of measurement in the predictor and criteria were not amended for. The important finding of this study was that GRE-A was found to be a strong predictor of GGPA just in the same as GRE-Q and GRE-V were. Interestingly, in some cases, the GRE-A predicted the GGPA scores better than the other two sections of the GRE did.

Young, Klieger, Bochenek, Li, and Cline (2014) probed the validity the GRE General Test scores in the admission of students in Master of Business Administration (MBA) program. This research gathered data from 480 students studying MBA. Concerning the GPA scores obtained during the first term, GRE-Q illustrated the most predictive effect followed by GRE-V and GRE-A, respectively. The GRE-A was also a weak predictor of UGPA. Further analyses showed that the three sections of the GRE test provided a significantly more predictive value than using only UGPA scores.

Although the international institutions in the USA are widely using the GRE revised General Test scores to select applicants for admission purposes in university contexts, we know little about the application of the test as an admission criterion particularly in other parts of the world. Therefore, our understanding about the analysis of the predictive validity of the GRE test outside the USA is apparently limited. Since the revised version of the General GRE test was released in August 2011, there is a pressing need to examine the predictive validity of the test, especially in contexts other than the USA (Klieger et al., 2016). The following two examples show the trend followed in countries other than the USA.

Schwager, Hülshager, and Lang (2014) and Schwager, Hülshager, Lang, and Bridgeman (2014) in two unpublished papers, investigated the predictive validity of the GRE test in terms of 282 international students' GPA scores and their success in finishing the educational programs they had enrolled in at a university in the Netherlands. The analyses indicated that the three sections of the GRE test exhibited significantly predicted the scores obtained on GGPA. Moreover, GRE-A was found to be a significant predictor of the scores obtained on the M.A thesis upon completion. The regression analyses revealed that altogether, the three sections of GRE were better predictors of students' GGPA than their UGPA, socioeconomic conditions, and language proficiency levels. Additionally, the three sections of the GRE were better predictors of students' scores on M.A. thesis compared to their English proficiency levels, UGPA scores, and socioeconomic conditions. Finally, the results did not

find support for the GRE scores and its three sections as significant predictors of students' success in finishing the educational programs and completing their M.A. theses.

Liu et al. (2016) investigated the predictive validity of the GRE revised General Test among master's and PhD students in a Singaporean context. The study also conducted interview sessions with chief admission committee members to know how the GRE revised General Test and its sections were being used in the admission decisions. The results indicated that GRE revised General Test scores significantly predicted the students' GGPA scores. Also, the GRE scores were found to be better predictors of students' submission compared to the GGPA and UGPA scores. Additionally, the students who had submitted GRE certificate in their submissions obtained better GGPA scores compared with the applicants who had not submitted their GRE scores.

Klieger, Bridgeman, Tannenbaum, Cline, and Olivera-Aguilar (2018) explored the predictive validity of the General GRE test with 1,587 graduate law students. Accordingly, they concluded that the GRE was a robust, valid, and reliable predictor of students' scores during their first year at law schools even when undergraduate GPA scores were taken into account.

Sealy, Saunders, Blume, and Chalkley (2019) investigated the association between the GRE scores and biomedical sciences PhD students' academic success. Taking the biased factors into account, they concluded that the GRE scores are weak predictors of students' academic success in the future.

Finally, Petersen, Erenrich, Levine, Vigoreaux, and Gile (2018) took GRE test scores as a predictor of PhD completion among a sample of 1805 doctoral students in four educational institutes in the USA. The results indicated that the GRE scores failed to predict students PhD program completion.

3. Methodology

This study is an analytical critical review of the predictive validity of the General GRE test. Based on the existing empirical investigations available in the review of the literature mentioned above, this study intends to answer the following research question:

RQ: Do the General GRE test scores predict students' academic success in future?

4. Discussion and Conclusion

Reviewing the above-mentioned studies, the authors of this study have prepared a critical commentary on the predictive validity of the GRE General Test. The main challenges discussed are criteria definition, the delayed effect of the test through time, and SES factor effect.

4.1. GRE tests only predict short-term restricted criteria

Dealing with the predictive validity, tests would not possibly claim to measure all important criteria for success. Ostensibly, the studies conducted up to the present time have examined myriads of criteria for success (e.g., GGPA, number of publications, comprehensive examination grades, ratings bestowed by departments, number of publication citations, degree attainment, etc.) (Kuncel et al., 2001), yet all these criteria seem to be just one piece of a greater dynamic puzzle. While the results have been promising at times, there are no empirical

researches, to my knowledge, showing the GRE test scores predict the delayed effect in the long run and in other areas such as career success. Nevertheless, an extensive amount of research contradicts this idea. According to [Sackett, Borneman, and Connelly \(2008\)](#), the meta-analyses (e.g., [Kuncel et al., 2001](#)) and large-scale samples investigations (e.g., [Kuncel & Hezlett, 2007](#); [Ling, Buzick, & Belur, 2020](#)) strongly suggest that the frame of reference regarding the efficiency and relevance of the GRE scores in decision making processes is not solely restricted to students' performance in their first year of education. [Sackett et al. \(2008\)](#) maintain the evaluation of the GRE scores could stretch beyond short-term goals and potentially predict various long-term academic achievements and even later career successes. While acknowledging the findings of studies of this kind, the researchers of this study contend that such findings explain the predictive validity of the criteria restricted to the academic success. Other important criteria such as employment setting conditions remain in the dark. Career success has not been the focus of the mentioned studies at all. Furthermore, test information considering what a university students could recall (e.g., comprehensive examination scores) may not be a good predictor of how the students will really perform in a real-life academic situation where critical verbal and analytical orchestration of thinking abilities are demanded. In academic contexts, students should exhibit creative use of new information, management of difficult research tasks, and the exploration of the unknown areas of knowledge ([Lerdau & Avery, 2007](#); [Razmi, Jabbari, & Fazilatfar, 2020](#)).

Additionally, completing the doctoral dissertation (degree attainment criterion) demands financial support, social and communicative skills, persuasiveness, perseverance, creativity, resilience, management skills, interest, independence, family integrity, physical and emotional health, and even luck. A relevant question can be posed here: does relying merely on GRE scores predict individuals' performance affected by all the factors mentioned? The other phenomenon related to the discussion is compensatory selection. Compensatory selection is a particularly used in admission decisions when candidates who have obtained low scores on the GRE may still be admitted if they could prove competency in other related application requirement such as interest, competence in conducting research, and high GPA scores ([Powers, 2004](#)). Consequently, a group of applicants with high scores on the GRE test may not find their way through the admission because they may lag behind in the evaluations of interest, research expertise, or any other application requirements. Conversely, some applicants with lower GRE scores may win the admission just because they have exhibited acceptable competencies in other admission requirements. [Ruscio \(1998\)](#), in this regard, maintains that "compensatory selection often stacks the deck squarely against the predictive validity of the GRE" (p. 569).

The criterion problem remains a challenging issue pertinent to the predictive validity studies. The ETS has launched a number of studies to study the validity of the GRE. Sadly, this field of study, like other areas of investigation, comes with its own limitations. According to [Goldberg and Alliger \(1992\)](#), researchers are seeking the answer to the question they fail to adequately define. The authors of the present investigation share the common ground with [Goldberg and Alliger](#) who maintain that investigators exploring the validity of GRE had better

start off from the very beginning at ground zero and define the criteria which adequately demonstrate what exactly the GRE is to predict.

One may argue that there are universities and job recruitment institutions which endorse established selection criteria that are strongly predicted by standard tests such as GRE; however, predicting diverse criterion variables is hard to be predicted by merely GRE scores. This somehow explains why universities generally make use of letters of recommendation and statements of purpose along with the criteria predicted by test scores. Therefore, overreliance on GRE scores and GRE predictive validity does not seem to be a reliable phase of decision-making process.

4.2. GRE fails to predict socioeconomic factor effects

Tests are simply a representation of wealth. In other words, they display socioeconomic status (SES) rather than developed skills. Not only does SES affect the participants' performance on the test, but also it determines their success/failure over the future decisions (Schwager et al., 2015).

Related factors to SES are coaching (Powers, 1985), bias against racial and ethnic minority groups, and gender (House, 1994). These socio(economic) effects are serious threats to the predictive validity of the tests. Group membership gives privilege to some examinees over others. Such influences need to be controlled for if a true picture of one's ability performance is to be drawn. According to Messick (1981), coaching is surrounded by test score effects. Coaching and special preparation on tests would leave a bad mark on construct validity and impair the predictive validity of a test (Motallebzadeh & Baghaee Moghaddam, 2011; Powers, 1985; Ravand & Firoozi, 2016). It is worth noting that despite my assertion on coaching, conclusions regarding the impact of coaching have been conflicting and somewhat equivocal.

There are interesting quotations regarding the predictive validity of Scholastic Assessment Test (SAT). For example, Lani Guinier, a law professor at Harvard (undated), claimed that "in the interest of truth in advertising, the SAT should simply be called a wealth test" (as cited in Zwick, 2002, p. 311). Similarly, Colvin (1997) stated that the "only thing the SAT predicts well now is socioeconomic status" (p. B2, as cited in Sackett et al., 2008). Kohn (2001) claimed that "the SAT merely measures the size of students' houses" (p. B12, as cited in Sackett et al., 2008). Such assertions show how socioeconomic factors can play a significant role in predicting students' success. The plausible question, here, is: does GRE account for such socioeconomic factors contributing to ones' success or failure? Many individuals are deprived of deploying their potential fully, only because of socioeconomic issues.

The major limitations reported on the estimation of 'true' predictive validity of GRE, does not imply that decisions based on GRE scores are inaccurate. What the authors of this study want to put forth is that if scores are to predict success, many intervening variables should be controlled for. There are good tests used inappropriately (Sackett et al., 2008). GRE can be regarded as a good test serving various sound purposes, yet it may not always be a valid choice, especially with regard to the criteria we expect the GRE scores to predict.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article

References

- Amirian, S., Alavi, S., & Fidalgo, A. (2014). Detecting gender DIF with an English proficiency test in EFL context. *International Journal of Language Testing*, 4(2), 187-203.
- Burton, N. W., & Wang, M. (2005). Predicting long-term success in graduate school: A collaborative validity study. *GRE Board Research Report*, 99(14), 1-61.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity for psychological tests. *Psychological Bulletin*, 52, 281-302.
- Educational Testing Service. (2012). *GRE information and registration bulletin*. Princeton, NJ.
- Goldberg, E. L., & Alliger, G. M. (1992). Assessing the validity of the GRE for students in psychology: A validity generalization approach. *Educational and Psychological Measurement*, 52, 1019-1027.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. London and New York: Routledge.
- House, J. D. (1994). Gender differences in prediction of grade performance from graduate record examination scores, *The Journal of Psychology*, 128(6), 695-697.
- Huitema, B. E., & Stein, C. R. (1993) Validity of the GRE without restriction of range. *Psychological Reports*, 72, 123-127.
- Klieger, D. M., Bridgeman, B., Tannenbaum, R. J., & Cline, F. A. (2016). The validity of GRE® scores for predicting academic performance at the University of Hawaii William S. Richardson School of Law. *Educational Testing Service*, 1, 1-19.
- Klieger, D. M., Bridgeman, B., Tannenbaum, R. J., Cline, F. A., & Olivera-Aguilar, M. (2018). *The validity of GRE® General Test scores for predicting academic performance at U.S. law schools* (Research Report No. RR-18-26). Princeton, NJ: Educational Testing Service.
- Klieger, D. M., Cline, F. A., Holtzman, S. L., Minsky, J. L., & Lorenz, F. (2014). *New perspectives on the validity of the GRE General Test for predicting graduate school grades* (GRE Board Research Report No. 14-03). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/ets2.12026>
- Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, 315, 1080–1081.
- 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.
- Kuncel, N. R., Wee, S., Serafin, L., & Hezlett, S. A. (2010). The validity of the Graduate Record Examination for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement*, 70, 340–352.
- Lerdau, M., & Avery, C. (2007). The utility of standardized tests. *Science*, 316(5832), 1694-1697.

- Ling, G., Buzick, H., & Belur, V. (2020). The GRE and students with disabilities: A validity study at 10 universities. *Exceptional Children, 86*(2), 193–208. doi:10.1177/0014402919872688
- Linn, R. L., & Dunbar, S. B. (1982). Predictive validity of admissions measures: Corrections for selection on several variables. *Journal of College Student Personnel, 23*, 222–226.
- Liu, O. L., Klieger, D. M., Bochenek, J. L., Holtzman, S. L., & Xu, J. (2016). An investigation of the use and predictive validity of Scores from the GRE® revised general test in a Singaporean university. *ETS Research Report Series, 1* (1), 1-24. doi:10.1002/ets2.12095.
- Messick, S. (1981). The controversy over coaching: Issues of effectiveness and equity. In B. F. Green (Ed.). *Issues in testing: Coaching, disclosure, and ethnic bias* (PP. 21-53). San Francisco: Jossey-Bass.
- McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79*, 493–505.
- Motallebzadeh, K., & Baghaee Moghaddam, P. (2011). Models of language proficiency: A reflection on the construct of language ability. *International Journal of Language Testing, 1*(1), 42-48.
- Norcross, J. C., Hanych, J. M., & Terranova, R. D. (1996). Graduate study in psychology. *American Psychologist, 51*, 631–643.
- Oldfield, K., & Hutchinson, J. R. (1997). Predictive validity of the Graduate Record Examinations with and without range restrictions. *Psychological Reports, 81*, 211-220.
- Pearson, K. (1903). Mathematical contributions to the theory of evolution: XI. On the influence of natural selection on the variability and correlation of organs. *Transactions of the Royal Society, London, Series A* (200), 1–66.
- Petersen, S. L., Erenrich, E. S., Levine, D. L., Vigoreaux, J., & Gile, K. (2018). Multi-institutional study of GRE scores as predictors of STEM PhD degree completion: GRE gets a low mark. *PLoS ONE, 13*(10), 1-15.
- Powers, D. E. (1985). The effects of test preparation on the validity of a graduate admission test. *Applied Psychological Measurement, 9*, 179-190.
- Powers, D. E. (2004). Validity of Graduate Record Examinations (GRE) General Test scores for admissions to colleges of veterinary medicine. *Journal of Applied Psychology, 89*, 208–219.
- Ravand, H., Firoozi, T. (2016). Examining construct validity of the master's UEE using the Rasch model and the six aspects of the Messick's framework. *International Journal of Language Testing, 6*(1), 1-23.
- Razmi, M. H., Jabbari, A. A., & Fazilatfar, A. M. (2020). Perfectionism, self-efficacy components, and metacognitive listening strategy use: A multicategorical multiple mediation analysis. *Journal of Psycholinguistic Research, 49* (6), 1047-1065.
- Rock, J. L., & Adler, R. M. (2014). *A descriptive study of universities' use of GRE® General Test scores in awarding fellowships to first-year doctoral students* (Research Report No. RR-14-27). Princeton, NJ: Educational Testing Service. 10.1002/ets2.12027

- Ruscio, J. (1998). Gatekeeping, compensation, and fallibility. *American Psychologist*, *53*, 568–569.
- Sackett, P., R., Borneman, M. J., & Connelly, B. S. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist*, *63*(4), 215-227.
- Schmidt, F. L., & Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science*, *2*(1), 8–9.
- Schwager, I. T., Hülshager, U. R., Bridgeman, B., & Lang, J. W. B. (2015). Graduate student selection: Graduate record examination, socioeconomic status, and undergraduate grade point average as predictors of study success in a western European university. *International Journal of Selection and Assessment*, *23*(1), 71-79.
- Schwager, I., Hülshager, U., & Lang, J. (2014). *Validity of the GRE® General Test and ETS® PPI in a European context*. (Unpublished manuscript).
- Schwager, I., Hülshager, U., Lang, J., & Bridgeman, B. (2014). *Improving fairness in graduate student selection: GRE, socioeconomic status, and undergraduate grade point average as predictors of study success in a Western European university*. (Unpublished manuscript).
- Sealy, L., Saunders, C., Blume, J., & Chalkley, R. (2019). The GRE over the entire range of scores lacks predictive ability for PhD outcomes in the biomedical sciences. *PLoS ONE*, *14*(3), 1-17.
- Walsh, M. M., Arslan, B., & Finn, B. (2021). Computational cognitive modeling of human calibration and validity response scoring for the Graduate Record Examinations (GRE). *Journal of Applied Research in Memory and Cognition*, *10*(1), 143-154. doi:10.1016/j.jarmac.2020.08.012
- Wilkinson, T., Shugart, A., Williams, A., & Riechel, M. K. (2021). Racial group differences in the predictive validity of undergraduate GPA, GRE, and MAT scores on counselor knowledge acquisition. *International Journal for the Advancement of Counselling*, *43*, 63–75. <https://doi.org/10.1007/s10447-020-09416-8>
- Young, J. W., Klieger, D., Bochenek, J., Li, C., & Cline, F. (2014). *The validity of scores from the GRE revised General Test for forecasting performance in business schools: Phase one* (GRE Board Research Report No. 14-01). Princeton, NJ: Educational Testing Service. doi:10.1002/ets2.12019
- Zwick, R. (2002). Is the SAT a “wealth test”? *Phi Delta Kappan*, *84*, 307–311.