

Steven Bakker, Iwa Mindadze, David Ziegler

External University Admission Testing: Assumed and Proven Validity

Abstract. *In view of the high-stakes character of university admission procedures the validity of instruments used is crucial in avoiding both false negatives (unjustly rejected) and false positives (unjustly admitted). Researching ‘predictive validity’ is a way to find out more about the false positives: do high scores on an admission test indeed promise high scores in university? Unfortunately this tells us nothing about the false negatives, the unjustly rejected, because they do not find themselves in higher education. How can we improve admission procedures to decrease both errors and increase fairness? The paper will present approaches and research aimed at increasing the validity of admission tests, and present approaches from The Netherlands, Georgia and Israel. As International Consultants for the Ukrainian Standardized External Testing Initiative (USETI) the authors are advising the Ukrainian Centre for Education Quality Assessment in its development of admission testing and with this paper wish to encourage the introduction of validity research in Ukraine.*

Keywords: *admission, testing, validity.*

The High Stakes of Admission to Higher Education

The quality of higher education has become a factor of utmost importance for the economy of any country. Any society experiences a great deal of pressure to increase the education level of its workforce and admit more students to higher education. A considerable amount of taxpayer money goes towards improving the quality of higher education and increasing the numbers of graduates. Countries design strategies to make their system of higher education more effective, and regions join forces to become more competitive. The Bologna Declaration of June 1999 aiming at establishing a European Higher Education Area is a typical example of this phenomenon. All stakeholders, including students, universities and employers feel and exert pressure on access to and the quality of the

education being provided. For individuals higher education is an important investment in potential future income earnings and appealing jobs. Employers request more workers with broader skills for their ever changing needs, and universities experience pressure to deliver and reduce the number of students leaving without certificates or diplomas. The first moment that the pressure is really felt is at the entrance gate to higher education. All stakeholders should benefit by admission procedures that are transparent and select those who can be trained to meet society's demands, but do not harm interests and waste human capital by rejecting students on unjust grounds. However, this is often not the case.

Threats to Fairness and Validity

In practice a lot of human capital and tax money is wasted in unfair and invalid selection procedures. Of course unfairness starts at the level of unequal access to educational facilities. Many students simply do not have access to an education of sufficient quality because it is not in close proximity to where they happen to live. Unfortunately, unfairness continues to be a factor in the selection for admissions process, because of invalid and sometimes even corrupt procedures. There are various ways to fight these threats. Access issues may be countered by admission on the basis of general skills rather than curricular knowledge and understanding. Corruption may be fought by taking the admission procedures out of the hands of those who abuse them. This is where in many former Soviet republics and socialist states external testing initiatives find their origin. And last but not least, considerable improvements in the quality of the instruments and procedures utilized may be achieved by conducting appropriate research with regards to their validity and discriminatory power. It is on these issues that this paper will focus.

Criticism on the Validity of Admission Tests

In several countries that have a longstanding tradition in using external testing as part of the university admission procedure, criticism has been voiced on the validity of tests that have been used. It is felt that 'something is lacking' in these tests that could be essential for success in an academic education.

In the US for instance the SAT came under fire. This test is an external assessment of general skills, and part of a procedure that also includes the assessment of subject skills, a letter of motivation and recommendations. Some, mostly prestigious, universities were unhappy with the SAT, deeming it insufficient because it was too general, and unrelated to curricula. The University of California

threatened to scrap the SAT in favour of a test more closely linked to high school coursework. Interestingly, it should be noted that the SAT was once introduced to increase equity by taking some distance from curricular knowledge.

In England, university admissions are based on a mix of internal and external subject tests (A-levels), and a letter of motivation. A general complaint is that these A levels are losing discriminatory power. More and more students are applying to university with top grades, and it is becoming more and more difficult to decide on the basis of these grades who should be given or denied access to the limited number of places for some popular programs of study.

In The Netherlands, where school diplomas give access to higher education, universities have been faced with increasing drop-out rates during the first year of academic studies. This raises the question of the predictive power of the existing procedure: do pass scores on internal and external tests used for awarding school diplomas promise success in future studies? Can this success be predicted by cognitive tests in the first place, or are there other, non-cognitive characteristics of students that should also be taken into account? And if so, how can we assess these? And if we can, how much do they add to the predictive quality of the admission procedure?

Validity

The increasing calls for reconsideration of the use of certain tests for admission purposes have emphasized the need for thorough evidence that support the validity of such tests, and the validity of any instruments that would be added to the procedure. 'Test Validity' is a concept that is not a property of a test itself, but only exists in relation to the purpose of the test. It refers to "the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Test validation is the process of accumulating evidence to support such inference." (American Educational Research Association/American Psychological Association/National Council on Measurement in Education, 1999). Such evidence may consist of comparisons of actual test scores with 'true' criterion scores, which may be scores from other, similar test available at the same time, or available at a later moment in the future of the test takers.

Several types of validity are distinguished in the literature on educational and psychological testing, depending on the availability of criterion scores. Construct validity, for instance, deals with the question whether a scale measures or correlates with the theorized psychological scientific construct (e.g., "fluid intelligence") that it purports to measure. Construct validity answers the question: "Are we

actually measuring (are these means a valid form for measuring) what (the construct) we think we are measuring?" If we believe that 'critical thinking' or 'problem solving abilities' are essential for success in academic education, and that these constructs are based on a well-established theory and can be operationalized in an educational test, then part of the validation of these tests should be collecting evidence that they are actually measuring the constructs of critical thinking and problem solving abilities. Dutch national examination tasks -which play an essential part in university admissions - for instance contain many tasks that are supposed to operationalise such constructs.

This paper will, however, focus on a different type of validity, the so-called predictive validity. The question to be answered here is: do the results on this test actually predict success in academia or the jobs for which it is used as an admission tool? For predictive validation of an admission or selection test comparison of the scores on the test to be validated will happen post hoc, with criterion scores that are measures of success in the field of study or job that the test was giving access to. For university admissions testing, for instance, a grade point average at the end of the freshman year could be used as a criterion score.

Two main questions will be addressed: how well can we estimate the predictive validity of cognitive tests, and how can we use these results for improving admission procedures? For a general overview to help us arrive at a better understanding of the predictive validation of admission tests a College Board report dealing with the predictive validity of the SAT is discussed. Then follow three case studies from The Netherlands, Georgia and Israel.

SAT and Predictive Validity

The SAT is the most widely used test in the US for university admissions. It is owned, published and developed by the College Board, a non-profit membership organisation that is comprised of more than 5,900 schools, colleges, universities and other educational organizations. The primary purpose of the SAT is to measure a student's potential for academic success in college. The test has three sections, a critical reading section (SAT-CR) a mathematics section (SAT-M) and a writing section (SAT-M). After the revision of the SAT in 2005 a predictive validity study was carried out on a large sample of students entering 110 four-year colleges and universities in fall 2006 and completing their first year of college in May/June 2007 (Kobrin et al., 2008). The study nicely illustrates the usual problems associated with predictive validation of admission testing:

1. To what extent does an admission test add anything to the information already available such as high school grades? It is often claimed that the high school grade point average (HSGPA), a composite of mainly subject test scores, is as good or even a better predictor than the SAT with its curriculum-unrelated general skills tests.

2. What is the measure for ‘success in academic studies’, which is the criterion to which the SAT scores should be correlated to in the predictive validation? Usually a measure of success at the end of the first year in college is taken, the first year grade point average (FYGPA). However, in different institutions the FYGPA can mean very different things. And as Kobrin et al. state: ‘Just as academic requirements and admissions practices vary across institutions, the relationships between standardized test scores and the FYGPA also vary across institutions.’

3. In predictive validation studies we often can only correlate scores of students who were admitted. After all, for those who were not admitted we have a test score but no criterion score, such as the FYGPA. This is the so-called ‘range restriction’ problem, which leads to an inherent underestimation of the predictive validity¹. The ‘raw’ correlation obtained for the restricted range may statistically be adjusted to get an estimation of the true correlation.

The study shows that the SAT and the HSGPA indeed are related, but that there is a far from perfect relationship between these two measurements (Table 1). This suggests that the SAT is measuring skills that are either not assessed or receive less assessment in high school tests, and vice versa. The assumption that the SAT does bring added value to the admission procedures is supported by the multiple correlations between the SAT and HSGPA with FYGPA (Table 2). While the correlation of the SAT (CR, M and W) alone with the FYGPA (.53) is similar to the correlation of the HSGPA with FYGPA (.54), the combination of the two brings the correlation with the FYGPA up to .62.

These figures are ‘pooled within-institution correlations’. Kobrin et al. first computed separate correlations for each institution (110 institutions took part) and then computed a set of average correlations, weighted by the size of the institution-specific sample.

¹ For a practical explanation of this phenomenon see <http://www.gmac.com/gmac/SchoolServices/VSS/UnderstandingValidity.htm> or <http://davidmlane.com/hyperstat/A68809.html>

In this procedure they also applied a statistical correction for range restriction to each set of correlations separately. In Table 2 both the unadjusted and adjusted values are shown.

The report concludes that the changes made to the SAT in 2005 did not substantially change how predictive the test is of first-year college performance. However, across all institutions the recently added critical writing section is the most predictive of the three SAT sections. As expected, the combination of HSGPA and SAT scores is still the best predictor of academic success, as measured by the FYGPA.

N=151,316	HSGP A	SAT-M	SAT- CR	SAT-W
HSGPA	--			
SAT-M	0.23 (0.49)	--		
SAT-CR	0.21 (0.45)	0.50 (0.72)	--	
SAT-W	0.21 (0.49)	0.50 (0.72)	0.71 (0.84)	--
SAT-M, SAT-CR, SAT-W	0.28 (0.53)			

Table 1 Raw (and Adjusted) Pooled Correlation Matrix of SAT and HGSPA.

From Kobrin et al, 2008.

Predictor(s); N=151,316	R aw R	A dj. R
1. HSGPA	0 .36	0 .54

2. SAT-CR	0 .29	0 .48
3. SAT-M	0 .26	0 .47
4. SAT-W	0 .33	0 .51
5. SAT-M, SAT-CR	0 .32	0 .51
6. HSGPA, SAT-M, SAT-CR	0 .44	0 .61
7. SAT-CR, SAT-M, SAT-W	0 .35	0 .53
8. HSGPA, SAT-CR, SAT-M, SAT-W	0 .46	0 .62

Table 2 Unadjusted and Adjusted Correlations of predictors with FYGPA.

From Kobrin et al, 2008

Adding non-cognitive measures: a Dutch experiment

Admission to higher education in The Netherlands is a special case in so far as school diplomas act both as a passport for leaving the secondary school system and as visa for entering the tertiary system. Diplomas are awarded on the basis of scores on school and national subject examinations. A diploma makes the holder eligible to university studies; universities are not allowed to run additional selections, except for a few special schools, mainly fine art and physical education colleges. They do, however, have the right to deny students access to the second year of studies if results in the first year do not meet a minimum criterion. The main reason that higher education institutions never felt the need for introducing separate admission test is the fact that already at the beginning of secondary education cohorts are separated into different streams, one specifically preparing for university and one

specifically for higher vocational studies. There is also a third stream leading to vocational training. Universities always relied on the pre-academic stream in the secondary system to give them students with the right skills and attitudes. National examinations reflect this by focusing on skills application and problem solving, and addressing competences such as communication, social contextualization and design and research skills.

Lately, two developments have put this system under pressure, though. One is the national policy to get more students to enter higher education. Currently this is about 30% of all students at age 18. The government wants to increase this number to 50%. This should not proceed at the expense of quality and efficiency: Dutch universities should be among the best in the world, ‘Centres of Excellence’ for highly talented students should be established and all students should complete their programs with no delays or transfers. Especially the latter is an issue: Dutch higher education institutions are being confronted with high drop-out rates during the first year of study.

The question is whether the system can effectively absorb the increase of students. While for some programs of study this is not a problem, a number have considerably fewer places than there are applicants. For these so called ‘numerous clausus’ studies (e.g. medicine, psychology, dentistry) it is a lottery that decides rather than a rank order of test results, although high test scores do give students the advantage of a higher weight of their lot. A number of cases of top students being denied access due to the lottery increased claims that this is ‘too much equity’ and that it is the cause of loss of talent.

Both developments, a broader range of applicant abilities and the wish to identify talent and nurture it, led to initiatives to introduce some tests as an entrance requirement alongside school diplomas. As it was known already that dropping out was caused by emotional and attitudinal factors rather than lack of cognitive skills, it was believed that the additional tests should set out to measure non-cognitive skills.

Predictor	GPA end of first year
------------------	----------------------------------

Average National Examinations	.50
IQ-test	.21
Motivation test	.29

In 2004 the University of Leiden got permission to introduce additional selection tests on an experimental basis. The university decided to investigate the effects of adding a cognitive IQ-type of test and a motivation test. In this experiment instead of rejecting or

accepting students on the basis of these tests, all students were allowed to enter and their test results were correlated with their FYGPA. Thereby not only the false positives would be identified but also the false negatives.

The results were rather disappointing. They showed that the results on national examinations (the basis for the school diplomas) were still the best predictors. The special cognitive selection tests did not improve the prediction, and the non-cognitive test (i.e. the motivation test) was neither reliable nor

Table 3 Correlations experimental entrance tests and valid. The latter was probably due to lack of motivation of students and the

fact that they could easily manipulate the test. Leiden University decided not to continue the experiment. The government concluded that in the Dutch situation ‘selection at the gate’ is generally not desirable, but rather that selection at the end of the first year is the best way to avoid later delays and dropping-out and also that identification of true talent for centres of excellence can better be done during the period in which a student is studying rather than at the start. To counter the drop-out rates in the first year the government is now discussing changes in the relationship between school and national exams, and putting higher demands to scores on certain key subjects such as Mother Tongue, Foreign Languages and Math.

Incomparability of Criterion Scores: a Georgian investigation

Georgia saw external independent admission testing for the first time in 2005, as part of an effort to fight the corruption that was riddling admissions to universities. Up until that time, getting your child a place in a university was very much a matter of finding a middleman who could put you in touch with the right coach, taking into account the means of the parents. These coaches usually were the same university professors who set the entrance exams. In many universities coaching was very

much part of the regular activities of the institution. The rector of Georgia's largest university, Tbilisi State, owned a bank account in which he accrued all coaching fees paid by students, and from which he transferred bonuses to his staff, keeping a nice percentage for himself (Bakker, 2012). The corrupt system had a detrimental effect on education. Most students in their final grade of secondary education did not attend classes anymore but rather relied on cramming facts indicated to them by their tutors as necessary to know for the test. And once admitted there was little reason for students to engage in their studies either. The next exam results had to be bought as well.

The new Minister of Education who came in after the Rose Revolution of November 2003 made it a priority to break the stranglehold of the 'academic mafia'. As a first move, he replaced the university-run admission by independent external testing and allocation of students to available places by a governmental body. This task was put in the hands of the National Assessment and Examination Centre (NAEC).

Next to tests of Georgian Language and Literature and Modern Foreign Languages, the NAEC was asked to design a general ability test consisting of a verbal and math section. This test was meant to give a fair chance to students who had not had access to an appropriate education but possessed the skills that would make them successful in university. As of 2006 natural and social science subject tests also became part of the procedure. The distribution of scholarships, however, would remain largely based on success in the general ability test.

The first administration of the national admission tests was a huge success, especially in the eyes of the public at large. Parents who could watch CCTV images of the testing process on screens erected outside the centres reacted with increasing enthusiasm. "Now we can see that our children are treated in a friendly and polite way, and not like cattle, as before. We see with our own eyes that they are seated properly, at a good desk on a proper chair, and that each is given a fair chance", a parent told the press. "I was expecting the same games that were played before, but this is totally different!" For them it was a relief not to be caught any longer in a system where everybody was cheating and playing along was the only option. Soon after the first test was administered, a large number of corrupt rectors were dismissed and replaced. It was clear, however, that in order to keep the trust of all stakeholders - students and their parents, schools, universities and politicians - the quality of the procedure and validity of instruments would have to be proven year after year.

As part of these efforts NAEC started a program of predictive validation. Over a period of three years (2005, 2006, 2007) it collected FYGPA data of those who had taken part in the admission tests

and correlated the data. As school results were considered to be unreliable and were known to show only little intra-school variation – some schools were only delivering candidates with top grades for example – comparisons with high school grade point averages were not made part of the study. Below, the results of these analyses of 2006 are discussed. The results from 2004 and 2005 are not dissimilar.

Predictor	Correlation FYGPA	
	I term	II term
Georgian Language and Literature	0.40	0,40
General Ability Test	0,40	0,40
Foreign Languages	0,55	0,51

Table 4 Correlation Georgian universities' FYGPAs with external admission tests

Table 4 shows the results of the correlation of a representative sample of students from all high schools in Georgia with their First Year Grade Point Averages at the end of the first and second term, after correction for range restriction. Cohen (1988) offers general criteria for interpreting correlation coefficients: an absolute value of around 0.1 means a weak correlation, around 0.3 means medium and 0.5 or higher a strong correlation. According to these rules of thumb the predictive validity of the external admission tests in Georgia are sufficient to strong, with little attenuation over the year. The latter may be expected as during the year items other than cognitive factors may come into play, co-determining students' academic successes.

A more detailed look, however, shows that this is not true for each and every university in Georgia. The 'criterion score problem', the fact that FYGPAs mean widely different things from one university to the other, manifests itself when correlations for different universities are calculated and compared. While being less reliable because of the smaller number of observations, the findings are nevertheless meaningful (Table 5).

University	Semes	Georg	GAT	Foreig
Ilia Chavchavadze State University	I	0.39	0.40	0.32
	II	0.36	0.36	0.27

State University of Agriculture	I	0.14	--0.17	0.03
	II	0.16	--0.05	0.07
Caucasian University (private university)	I	0.38	0.31	0.25
	II	0.30	0.17	0.23

Table 5 Correlation FGPA individual universities with external admission tests

The State University of Agriculture is well-known for its relatively light and selective assessments at the end of semesters, with almost all students coming out with the ‘best’ marks. The other two are more prestigious institutions with a reputation of rigorous assessment.

Israel: predictive validity studies for better admission procedures

Israel is a small, crowded, and fragmented country. Its population amounts to about 7.5 million of which approximately 6 million are Jews and 1.5 million are Arabs. Most of the latter are Muslims, but there are also Druze and Christian Arabs.

In general, admission to higher education in Israel is decided based on two scores: the Bagrut Diploma score and the score on the Psychometric Entrance Test for university. These two scores are used to calculate the composite admissions score (CAS). All high school students take the ‘Bagrut’, which is an array of internal and external examinations and evaluations on the basis of which the High School Diploma is awarded. A Bagrut Diploma is required by all institutions that grant an Israeli academic degree. The Bagrut Diploma score will from here forward be referred to as HSGPA.

The Psychometric Entrance Test is developed and administered by the National Institute for Testing and Evaluation (NITE). The PET is a general test of learning potential, similar in form and content to the SAT. Since 1990 the test has consisted of three sections: Verbal Reasoning, Quantitative Reasoning and English as a Foreign Language.

In most departments and schools, the CAS is used as a criterion for admissions. The procedure is meant to be purely meritocratic: blind to the applicant’s gender, religion, ethnicity, socio-economic status, etc.

From the moment the PET was introduced as a key criterion in the admission process for higher education, on the grounds that it is predictive of academic success, there has been public and political pressure to demonstrate this predictive ability. The PET is therefore probably the most researched test in Israel. This research has repeatedly found PET to be a good predictor of academic success in all

areas of study. But from time to time the doubts regarding the appropriateness and fairness of the university admissions process raise again questions such as: Is PET a better predictor than HSGPA? Is CAS a better predictor than both? Are all academic fields equally predicted by these measures? And above all: is there a way to use PET, HSGPA or any combination of them in a better way in terms of both prediction quality and fairness? Ongoing research into the predictive validity of PET and other tools used in university admissions is essential for trying to answer some of the above questions.

A study analyzing more than 56,000 students in different faculties, institutions and cohorts (2003-2005) showed that PET predicts both FYGPA and also, and to a no lesser degree, GPA at the end of undergraduate studies. This study demonstrated the stability of the predictive traits of the university admissions process. PET's predictive validity coefficient was 0.46, HSGPA had a predictive validity coefficient of 0.38 and the composite score, CAS, was once again found to have the highest predictive validity coefficient: 0.50 (Bronner et al., 2007).

	Hum anities	Law	Engin eering	Exac t/ Natural	Medi cine	All
Verbal	.47	.56	.39	.48	.44	.43
Quanti tative	.39	.53	.43	.48	.43	.38
Englis h	.42	.40	.35	.39	.31	.35
PET	.53	.62	.49	.65	.52	.46

Table 6 predictive validity of the PET and its different components across academic fields

As mentioned before, however, predictive validity is not only a value to aspire to reach; in many cases it indicates the direction in which to move in order to improve the admissions process. For example, the different subtests that comprise the PET (verbal, quantitative and English) each displayed different predictive properties in different fields of academic studies (Table 6). This finding raised the question whether PET subtests should be equally weighted or whether for the sake of slightly better prediction in some fields; it is recommended to weight them by field. For example the quantitative

subtest would be given more weight in the overall PET score for candidates who wish to study engineering.

Predictive Validity is also crucial when examining the fairness of a test. The fairness of the PET and the HSGPA with regard to various social groups has been studied extensively. The PET was found to be unbiased towards Arabic-speaking or Russian-speaking examinees. The PET was found to favour men while the HSGPA was found to favour women; the combined score, however, was found to be fair to both genders (Turvall et al., 2008).

The following recent study, which re-examined whether the higher education admissions process in Israel discriminates against the Arabic-speaking population, is an example of the role of predictive validity in identifying bias. The research included data for over 57,000 students from all universities in Israel who began their studies between 2004-2008. The specific measures examined were the FYGPA and six predictors, namely the HSGPA, the Psychometric Entrance Test (PET), its three subtests (Verbal, Quantitative and English) and the composite admission score (CAS). The question of fairness was analyzed from two points of view:

1. Differential validity – analysis of the degree of agreement between the predictors and the criterion across the two groups of examinees (Arabic-speaking and Hebrew-speaking)
2. Bias in selection (according to 3 different models) – under - or over - prediction of the criterion (FYPGA) as a result of using the same predictor regression line for both groups.

It was found that the validity of the predictors is considerably higher among the Hebrew-speaking examinees than it is among the Arabic-speaking examinees. But it was also found that the CAS – the predictor commonly used to screen the candidates – is biased in favour of Arabic-speaking students (Kennet-Cohen et al., 2011). This bias results from the fact that the HSGPA is highly biased in their favour and PET is either unbiased or slightly biased in their favour (depending on which of the 3 models is used).

Fairness studies conducted in the USA regarding the SAT (Mattern et al., 2008) showed similar results: lower predictive validity for minority groups (Hispanic and African Americans) as well as bias of all predictors in favour of these groups.

One way to explain the results in Israel might be the fact that the Arab students are tested in Arabic for determining the predictor (PET) but the criterion is measured based on university teaching and testing in Hebrew. This explanation is supported by the fact that in more verbal academic fields the

bias in favour of Arabic-speaking examinees is bigger - in other words the more the criterion is verbally dominant, the more the predictor is biased in favour of these examinees.

An alternative explanation might be considered, based on parallels with the SAT research: the minority's potential is actually measured correctly but they fail to fulfil their potential due to prejudice or economic hardship encountered during their studies.

Conclusions

Predictive validation of admission tests is a 'conditio sine qua non'. Without proper data that support the predictive power of a test it is irresponsible to use it to give or deny students access to higher education, from an ethical and fairness perspective but also from an economic and political perspective. Unreliable instruments will lead to a waste of talent and undermine the confidence of the public in admission procedures, which again may jeopardize efforts to take these out of the sphere of corruption and bribery.

Well-founded research based on predictive validation data may yield invaluable suggestions for improving admission procedures. It may answer questions about weighing components of existing tests for better decision-making, and the use of additional cognitive and non-cognitive demands.

Unfortunately, predictive validity of admission testing often is regarded by the public, politicians and the press as the "bottom line" of a test: most people are uninformed in psychometric concepts and principles. Not uncommon, even among university professors, are claims such as: if the predictive validity of a test is 0.5 it means that the quality of prediction is equal to "coin flipping" chance, or in other words, 50% of the cases are errors of prediction. The public at large, however, repeatedly proves to be intolerant of errors of prediction. Just as it is difficult to accept the idea of measurement error, it is also difficult to accept the idea that even the best predictors err, and that prediction errors (false negatives and false positives) are inseparable from correct admissions and rejections. Thus when presenting predictive validity studies it is always worthwhile to explain that prediction involves errors of prediction, but that the existence of errors does not invalidate the tool. The important question is not whether there are errors, but rather how many errors there are, how large they are, whether they correlate to certain relevant variables and finally, is there a better alternative - a test which will prove to be more just and better predictive.

References:

1. American Educational Research Association/American Psychological Association/National Council on Measurement in Education (1999), *Standards for Educational and Psychological Testing*.
2. Kobrin, Jennifer L, Patterson, Brian F., Shaw, Emily J., Mattern, Krista D., and Barbuti, Sandra M., (2008), *Validity of the SAT[®] for predicting First-Year College Grade Point Average*, The College Board, New York, NY, USA
3. Bakker, S.A. (2012), *Introduction of External, Independent Testing in 'New Countries'. Successes and Defeats of the Introduction of Modern Educational Assessment Techniques in Former Soviet and Socialist Countries*, Educational measurement: Issues and Practice, 31,2 (in print)
4. Cohen J. (1988) *Statistical power analysis for the behavioural sciences*, Hillsdale, NJ: Lawrence Erlbaum Associates.
5. Bronner, S., Kennet-Cohen T., and Oren, C. (2007). *Accumulated data on predictive validity of PET*, Report No. 342, National Institute for Testing & Evaluation, Jerusalem, Israel.
6. Turvall E., Bronner, S., Kennet-Cohen T., and Oren, C. (2008). *Fairness in the higher education admissions procedure: the PET in Arabic*, Report No. 349, National Institute for Testing & Evaluation, Jerusalem, Israel.
7. Kennet-Cohen T., Oren, C., and Turvall E. (2011). *Fairness in the higher education admissions procedure: the case of gender and Arabic And Hebrew PET versions*, Report No. 376, National Institute for Testing & Evaluation, Jerusalem, Israel.
8. Mattern, K. D., Patterson, B. F., Shaw, E. J., Korbin, J.L., & Barbuti, S.M. (2008). *Differential validity of the SAT*, SAT College Research Report No. 2008-4, New York: USA.
9. Cohen, Y. *Issues in Educational Measurement and in University Admissions – The Israeli Scene*. National Institute for Testing and Evaluation, Jerusalem, Israel. (as presented in Astana conference , 2012).