Grand Design Instrument of Potential Abilities for Placement Test

Sutrisno

Department of Education of Islamic Science, UIN Sunan Kalijaga, Yogyakarta, Indonesia
trisno_63@yahoo.com

Asih Widi Wisudawati

Department of Chemistry Education, UIN Sunan Kalijaga, Yogyakarta, Indonesia and Chemistry Education Institute at Muenster University, Germany
a_wisu01@uni-muenster.de

Abstract

This research aims to develop potential ability instruments for placement tests in graduate degrees. This study used the research and development instrument made by Hambleton and Jones sequences. The potential abilities measured by the instrument are academic potential and language skills. The product was reviewed by professional assessor and used in admission for selection in the academic years 2016/2017. The professional assessor reviewed the content, test construction, and language aspect. The construct and content validity from the professional assessor reached a good level with minor revision for the next selection. Ultimately, the results of the reviewing process are the potential academic test must not measure prior knowledge but the initial potential of the candidate and test for language should be emphasized on prior knowledge for Islamic study.

Keywords: *graduate placement test, instrument, potential ability, test items*

Abstrak

Penelitian ini bertujuan untuk mengembangkan instrumen kemampuan potensial untuk tes penempatan di tingkat pascasarjana. Penelitian ini menggunakan instrumen penelitian dan pengembangan yang dibuat oleh Hambleton dan Jones. Kemampuan potensial yang diukur oleh instrumen adalah potensi akademik dan keterampilan bahasa. Produk ini ditinjau oleh penilai profesional dan digunakan dalam penerimaan untuk seleksi di tahun akademik 2016/2017. Penilai profesional meninjau aspek konten, konstruksi tes, dan bahasa. Konstruk dan validitas konten dari penilai profesional mencapai tingkat yang baik dengan revisi kecil untuk seleksi berikutnya. Pada akhirnya, hasil dari proses peninjauan adalah tes akademik potensial tidak boleh mengukur pengetahuan sebelumnya tetapi potensi awal kandidat dan tes bahasa harus ditekankan pada pengetahuan sebelumnya untuk studi Islam.

A. Introduction

The primary purpose of a college is to give the best education for students to prepare for their future, also will create a future generation in developing their academic research and the evolution of students' career paths (Laurillard 2007; Nilson 2010; Jehanzeb, K., & Bashir, N.A 2013; Collins 2017). An indicator of the quality of teaching process can be achieved if there are abilities and willingness of the students to participate in the teaching process (Nilson 2010; Joe et al. 2017; Khajavy et al. 2018), which is conducted in Sunan Kalijaga State Islamic University Yogyakarta as the oldest of the State Islamic University in Indonesia. Student participation can become feedback for improving the instruction procedure. One of the quality factors of end products (graduates) is determined by the character of input (prospective students) (Mulyasa 2016). To ensure the quality of student input, prospective students must be taken through a credible, valid, and reliable selection system.

The development of the State Islamic University, like UIN Sunan Kalijaga Yogyakarta toward a better higher education organization, delivers three primary characteristics: improving quality and relevance, broadening access and fairness, and broadening autonomy. This concept involves the necessity of the State Islamic University to improve the quality of prospective students so the university could filter the best students who ready to become global citizens that should have the rigid potential ability and international language skills. Moreover, the English

course should accomplish and adjust standard competencies (Mujab 2016). In fact, that course has applied several methods (Erzad 2016) but the results from OECD, Indonesia remain has a lower grade compared to OECD countries (OECD, 2018)

As a promising Islamic State University, the university that organizes further studies after the undergraduate program should be capable of having high achieving student candidates and predicting the success in completing advanced work in this higher education timely and with a very satisfactory achievement. On the foundation of such lines, the new admissions test equipment is planned to measure common criteria that influence the success of prospective students in all courses or concentrations (Sithole et al. 2016; Noaman et al. 2017).

Quality of intake students is as a guarantee of improving higher teaching (Laurillard 2007; Nilson 2010; Brewis 2019; Jaiswal et al. 2019). The quality of intake students is measured by an appropriate placement test (Kern et al. 2015). The goal of this placement test is to assess the abilities of candidates who register at UIN Sunan Kalijaga Yogyakarta. Prior knowledge and potential of student candidates must be assisted with standard instruments to predict the outcome of higher education and to prepare a person who has 21st-century skill and career readiness. The principle of the selection process, however, remains to consider the ten targets of Sustainable Development Goal 4 that give equal access to high education to candidates (https://unesdoc.unesco.org/).

The instrument of measurement in selection purposes is required to map potential abilities such as verbal, nominal, spatial, reasoning, and language skills like Arabic and English, which are related to intelligence. An instrument is a tool which is needed to measure natural or social phenomenon including individual characteristic or the potential ability of a person (DeVellis 2012). The instrument is valid if it can measure what is to be measured, while the instrument reliable if it has a level of accuracy and consistency of the measurement results (Mertens 2015). The instrument is a measurement that has the ability to uncover the facts into data. The test is an objective measurement of a sample and standards of behavior (Privitera 2016; Kline 2015). While, the test is a systematic procedure to measure a person's behavior that is considered sample objectivity, standards and quality requirements of other tests (Kline 2015; Gravetter and Wallnau 2017). The interest in the measurement is to determine the quality of a person.

Supervision of new enrolment policies, including the development of the instrument used, greatly affects outcomes (Kline 2015; Derrington and Campbell 2015). Good quality student input guarantees should be selected through a credible, valid, and reliable selection system, keeping in mind the real needs of an institution and considering prospective students with high academic achievement and predictably successful completion of advanced studies in higher education (to predict the future success). Based on these arguments, the new student admissions test instrument is designed to measure common criteria that determine the success of prospective students in all courses and concentrations. The skills one has acquired are required for academic success in college (Wiliam 2011).

Measurement of cognitive abilities contributes to academic success, including average Grade Point Average (GPA) and cognitive acceptance tests for new students (Kyle, V.A., White, K.M., Hyde, M.K., Occhipinti, S. 2014). The selection patterns for written examination using new student selection instruments consist of Academic Potential Test, Arabic Proficiency Test, and English Proficiency Test. The Academic Potential Test is aimed at measuring the potential abilities that are possessed by a person, while Arabic and English proficiency test is a measure of cognitive abilities that one has acquired.

Potential abilities as assumed attributes to be a form of latent ability are not yet apparent from performance (Azwar 2009; Monteiro et al. 2015). The developed measuring tools such as Social-Relational Support for Education (SRSE) can be used to view latent capabilities, influenced by external factors that support parents in education, teacher attention in school, the suitability of values, sense of belonging in school, and friends at school (Vickers, M., Finger, L., Barker, K., Bodkin-Andrews, G. 2014). Potential abilities are formed from within oneself and influenced by external factors. The ability measured by the Academic Potential Test can be predicted by how far the ability of prospective students in facing academic problems. Appropriate measurement of potential abilities will be able to determine the quality of the education process at the postgraduate. The fixed test instruments will influence the right decision making in recruiting students. The selection process of admission of UIN Sunan Kalijaga Yogyakarta for graduate-level requires the use of standard test instruments in which had not undertaken yet by the scientific approach before. The standard test instances must be applied in a measurement process. Measurements include the process

of description and quantification (Keeves and Alagumalai 2005). The basic skills to be measured are the academic potential of new student candidates, reading, and writing ability in Arabic and English.

The test instruments are developed for placement tests during the admission process are meant to measure the abilities such as verbal, numeric, reasoning, spatial abilities, comprehension of texts in Arabic and English, the ability to write, and to translate from Arabic and English into the Indonesian language. If the prospective students can achieve these abilities, they are able to be estimated to perform good academic performances at the graduate level. As the standard instrument of new graduate admissions, the SAT (Scholastic Assessment Test) I and II are designed to assess the knowledge on certain objects (verbal, mathematical, literary, chemistry, and world history) and to predict students' readiness to undertake university-level learning (Keeves and Alagumalai 2005)

The most well-known test type in the graduate admission processes is the Graduate Record Examination (GRE) that has been used in more than 160 countries in 2017 in more than 1000 test centers worldwide (www.ets.org). GRE is the most widely used test in graduate student admissions, with GRE, the best criteria appear to be graded in the first year of the graduate program (Kaplan and Saccuzzo 2013). However, such tests like GRE cannot predict the clinical ability or problem-solving abilities in the real world (Neisser et al. 1996). Part of the GRE is the verbal capability (GRE-V), quantitative (GRE-Q), and analytical reasoning (GRE-A) in the form of multiple-choice options, which further development becomes a matter of description/essay. The verbal section includes reasoning, identification of the opposite, the use of analogy, and paragraph understanding while the quantitative part includes reasoning arithmetic, algebra, and geometrics.

Standardized tests, such as GRE and SAT, help the admission of a university or a college to decide to accept prospective students based on pre-requisite for aptitude tests designed to predict a person's ability at lectures and achievement tests capable of predicting lecture-performance (Slavin 2011). The standard test through the stages of detailed instrument development will be able to distinguish one's initial ability (prior knowledge). Every higher educational institution in today's world has a different learning outcome, and this indeed occurs in UIN Sunan Kalijaga Yogyakarta. Ultimately, the admission office at UIN Sunan Kalijaga needs to develop an appropriate selection test to recruit

the potential prospective student as GRE and SAT deal with, however, remain considering the own learning outcome.

The purpose of measuring a student's prior ability to be accepted is to predict the performance of a person. Language ability is the accumulation of language concepts that have been studied by someone, while the potential test of academics is the potential of a person associated with the pattern of completing a task or problem in the process of study. Therefore, the problems studied in this study can be formulated in the form of research questions as follows: How do the characteristics of the process and product of used instruments in the admission process in postgraduate selection at Sunan Kalijaga State Islamic University?

B. Material and Methods

This type of research is research and development (R & D). The intended instruments of selection test used in admission is an object of this research. Whereas the subject of this research is 1010 participants enrolled in the selection process at academic years 2016/2017 and registered participants in the same academic years around 519 people. Stages of development of selection test instruments according to Hambleton and Jones (Hambleton and Jones 1993) namely:

1) Phase preparation of test specification.

This first stage had done by collecting the problem by compilation team in Focus Group Discussion (FGD) to compose the latticework of questions to be developed and the minimal potential abilities that must be owned by the selected participants to be eligible for higher education. Academic potentials that be measured in this research include verbal, numerical, reasoning, and spatial dimensions. The ability to read and translate is a dimension in the English and Arabic skills test.

2) Phase preparation of the test item.

This stage of the problem-making teams begins to develop questions based on the Focus Group Discussion on the preparation of the latticework of potential abilities. After that, the items are prepared based on the latticework. A review process of the items has been developed. The review process will be done by the ways of questions that have been collected to the committee and then given to the expert judges to analyze the content and construct validity.

3) Field testing

Trials are conducted on the admission selection tests in the academic years 2016/2017. The results of the experiments were analyzed using ANATES to assess reliability, distinguishing power, and difficulties level.

4) Revision of the test Items

At this stage of the review process by professional judgment is different from the first review. The review process also includes the results of the analysis of trials and quality items test instruments covering material aspects, items test construction, and linguistics. The results of the review are used as the basis for the problem improvement process to be used in next year's selection. At the end of this stage, the reviewer performs an instrument feasibility assessment.

Assessment of the feasibility of the instrument is performed by the reviewer using a questionnaire in the form of a questionnaire. At this stage also performed data analysis to obtain the quality of the instrument with the steps as follows:

1. The obtained data are tabulated for each component and the subcomponent of the items test of assessment available in the instrument is the problem construction factor covering the material aspects, construction aspects, and linguistic aspects.

2. After collected data, then counted average score with the formula (1):

$$\overline{X} = \frac{\Sigma X}{n} \qquad ….. (1)$$

Notification:

$\overline{X}$ =   average score

ΣX =   total score

n   =   total items

3. To change the average score into a qualitative value with ideals measurement criteria as follows (Sukarjo, & Sari, L.P. 2008) that can be seen in Table 1.

Table 1. Ideals measurement criteria

No	Score range	category
1	Xi+ 1,8 SBi< X	excellent
2	Xi+ 0,6 SBi< X≤ Xi + 1,8 Sbi	Good

3	Xi - 0,6 SBi< X≤Xi+ 0,6 Sbi	Sufficient
4	Xi- 1,8 SBi< X≤ Xi- 0,6 Sbi	Milt
5	X≤ Xi - 1,8 Sbi	Very milt

C. Discussion

The results of the developed instrument are seen from the material aspect, item test construction, and linguistic aspects for Academic Potential Test, Arabic Proficiency Test, and English Proficiency Test. After that, both validation processes either content and construct validation are performed by professional assessor. The Academic Potential Test obtained a total of 58 scores including in the "B" or "Good" category including verbal, numerical, reasoning, and spatial dimensions. The Arabic Proficiency Test gets a total of 57 scores including the "B" or "Good" category. The English proficiency test gets a total score of 57 under the category "B" or "Good" level.

The analysis of the Arabic Proficiency Test shows the consistency of its reliability score on the first and second stage selection. The reliability at the first selection is 0.92, and the second selection is 0.82. The difficulties level from each item test is showed normal distribution at a moderate level. This level gets the most account than another level like easy, difficult and more difficult. Based on the result of ANATES software, the validity of the Arabic Proficiency test gets 38 items test was included by valid or significant.

The analysis of the Academic Potential Test describes the reliability of verbal, numerics, reasoning, and spatial ability are 0.51; 0.72; 0.65; and 0.51, respectively with a standard deviation between 2.80 to 3.45. The difficulties level from each skill test are reached proportional distribution, both more easy, easy, moderate, difficult until more difficult. The number of tests for each dimension of potential ability gets validity approximately 50% from the total number of items test, and more than 75% number of item tests can distinguish between participants with excellent skill and poor skill.

The result from validation by Professional Assessor for Instrument test has developed for the postgraduate level. The potential academic test question is a matter of exploring academic potential so that the questions are made as generic as possible so that those instruments do not influence by a matter of learning experiences. The landfill might contain many religious materials that probably can explore potential test-takers who

do not have religious material. However, there are several distractors can be an alternative option; the expert suggested that numeric item test should be completed with picture or graphic. The appearance of pictures or graphics can help test takers/testees to understand the test items. Academic Potential Test was included measurement of potential cognitive abilities a person, so must follow:

1. By design to reveal potential cognitive skills
2. Arranged based on prior abilities concept which was measured
3. Minimal related to learning memories or prior knowledge
4. A score of items test relevant to the probability of future success
5. Higher score achievement was obtained by general strategies to cope with a problem/ to solve a problem
6. Stressed in predictive validity

The appropriate academic potential test is developed using analytical power, logic, and reasoning of test-takers. The verbal test must assess in cognitive potential cluster analysis. After that, the numeric test must asses cognitive potential in logical thinking or why a person thinking about something. Whereas, in reasoning, test items must use logical reasoning and chart reasoning. The academic potential test is purposed for all people. So, all levels of enrollment use the same instrument in academic potential test.

The subsequent issue has discussed in Focus Group Discussion (FGD) is The Arabic Proficiency Test. The result from those discussions describes the appearance of test items in which it is more general. Moreover, the items test was not engaged with relevant knowledge or academic concentration in Islamic Study yet. From that instrument, experts have found grammatical aspect excess in two number of item test-number 20 and 21-from total items test. The model of question likely less variety particularly at the reading text that used modern text. The classical reading text perhaps difficult to find in the text. The other test is the English Proficiency test. The expert gave suggestions that the test items should include the competence to be measured. That competence should be an entry in the latticework.

Ultimately, the level of difficulty or level of competence should be described in the latticework. The appearance of validation is good, although it only looks at the physical aspects of the problems. Consequently, it has not seen prompted competence. In general,

the difficulty level of the problem is at a level that is high enough for prospective students.

Comparatively, selection processes for prospective students have been undertaken in this study. A qualitative data of enrollment from benchmarking with others admission have been used for a such comparison. Subsequently, the selection process and the selection instruments used in the enrollment of the Master Program from another university-Yogyakarta State University- have resulted the first information about enrollment mechanism, which describes two lines of admissions for master programs, namely the portfolio and paper test. The portfolio is for a selection of prospective participants from abroad while for participants from within the country is a written test. The written test is held in three waves with different problems. The second information about the specification of the instrument in writing test, which describes the use of the selection process of new admissions, is the Academic Potential Tests and English on clusters paper-based test (PBT). The test instrument was developed by a team of graduate students. The English test instrument has the equivalent difficulty level of TOEFL without listening.

The average result of academic achievement in GPA in the first year, which be compared for result in placement test with academic achievement in first years or GPA. The comparations present about the skill of a student in academic managerial as a part of cognitive individual problem-solving skills and social collaboration skills.

Table 2. GPA of Student in Post Graduate Sunan Kalijaga Islamic State University in the first semester

No.	Description	Result
1	A percentage of satisfactory cluster in GPA (up from 3.71 GPA Score)	60.5%
2	A percentage of the less satisfactory cluster in GPA (under from 3.71 GPA Score)	39.49%
3	Average of GPA achievement	3.68
4	Score minimum was achieved	0.42
5	Score maximum was achieved	4.00

1. The Important of an instrument in the selection process

The selection process is a process for determining or selecting a subject. The selection process for prospective students with a large number of applicants and limited quotas requires a reliable selection tool for screening all participants who entrants to the State Islamic University.

Instruments must have validity and reliability with comparable and economic results (Mardapi 2008).

Determination of the selected participants based on the achievement of scores will be obtained from the test instrument being worked on. The test results from the participant will be compared with the other participants, so that it will be know the position of a participant with other participants. This reference is used primarily in selection tests, following its purpose to differentiate a person's ability called a norm reference test.

The developed instrument is capable of identifying a dynamic latent potential that a student possesses. The instruments are capable of detecting a person's potential for study. The instruments are also capable of detecting foreign language mastery. The dynamic latent variables include attitude and personality patterns that change at any time (Li et al. 2016)

The purpose of a measurement instrument for the selection stage is to predict the potential or ability of a person to be predicted to be able to complete the tasks that must be completed in higher education. The developed test instrument is to measure the maximum potential. A person's potential can be used to predict whether a person can complete a duty, especially in the study in higher Institutions.

The internationally-used standard instrument GRE® can predict students' analytical abilities in the first year of study; a high correlation is between GRE® verbal proficiency scores and analytical abilities against student achievement outcomes in the first year (Enright and Powers 1990). The Potential Ability Test instrument that has been developed for this master program is expected to be able to measure the potentials of all participants by ignoring the background of the discipline they possess. However, the results of expert review/professional assessor obtained data that the Academic Potential Test still utilize the ability of religious materials and do not generic/general yet. In contrast to the Arabic proficiency test instrument, the Arabic Test instrument aimed at identifying participants' ability to understand classical texts and dealing with academics particularly for Islamic Studies. The data from professional assessor showed that the Arabic language tends to use modern/popular texts. Meanwhile, the English Proficiency test instrument aimed to measure the ability to translate and to understand the reading, such as the latticework in Table 3.

Table 3. Rated Aspects of Content Validity Process

No.	Rated Aspects
I.	Material aspects
1.	The suitability of statements with the formulation of indicators in the latticework
2.	The suitability of the measured aspect of each statement with the demands on the latticework
3.	There is only one right answer to every item>s test.
II.	Clarity and assignment of formulation of the principal
4.	The sentence is freedom from the negative statement or bias
5.	The sentence is freedom from statements referring to the past
6.	The sentence is freedom from a statement that can be interpreted as fact
7.	The sentence has no connection with the statements that can be interpreted in more than one way
8.	The sentence is freedom from statements that may be approved or vacated by all respondents
9.	Freedom of sentence from an uncertain statement (like all, always, sometimes, none, never)
10.	Each item has no relevance with other items
11.	The drawing (for Academic potential ability test) used helps participants understand the items test
III.	Linguistic aspects
12.	The rigor of the use of grammar in the statement
13.	Use of commonly used words/terms
14.	Communicative formulation statement sentence

2. Analysis of item test

A good instrument, in addition to have high validity and reliability, should be able to distinguish clever and fewer participants with high distinguishing power indexes, the difficulty level of proportional items according to the purpose of test preparation, correlation coefficient above the specified standard, and it has a good distractor quality. Arabic proficiency test instruments showed consistency scores on aspects of reliability, distinguishing power, and question validity. The reliability of the first and second test scores indicated the high category between 0.82-0.92. For differentiating power, the results of empirical validation indicated the number of items having a distinguishing power of> 0.30 is 35 points for first selection and 34 in the second selection. It indicated that the item can distinguish competent participants or not. The empirically obtained validity for the first and second selection is the same, i.e. 38

number of items test. It shows that the Arabic language test has a high consistency for two selections and has good quality for each item.

Instruments of potential academic tests were analyzed into four dimensions, namely verbal, numeric, reasoning, and spatial. The results of the empirical analysis of item points for the verbal dimension and spatial capability showed sufficient category reliability. The instrument of the academic potential test was effective to distinguish power, although the validity for each dimension could not be proved empirically. The potential academic instrument was developed based on a sequence of developing instruments that get proportional results and according to the expert shows a low-quality distractor and is seen from the analysis results per item.

The multiple-choice questions on the developed academic potential ability test and Arabic proficiency have met the minimum requirement as standard questions for the selection of new admissions because they have been able to differentiate which participants are who competent or not competent. The instruments would predict the ability for success study in the master program. However, the developed instrument needed to be revised to become a standard test in the selection process. Meanwhile, the English Proficiency was not analyzed per item by ANATES software because the type of test was the understanding of reading and translating with the corrector of more than five people.

According to professional assessor data, the test instrument was too general, and it had not correlated with the texts that correspond to the field of academic concentration, too many grammatical, less varied question models, and the reading texts used modern texts when the modern texts needed to combine with the classical text. Meanwhile, in the review of the English Proficiency should be included the competencies to be measured contained in the grid with the stated level of difficulty or level of competence to be measured by the university. Validation of the instruments performed in this research process was good but it only looked at the physical aspects of the problem, and have not seen the competence to be measured. In general, the difficulty level of the problem was at a level high enough for prospective students. The professional assessor data demonstrated the importance of validation and review processes for questions to be tested in the selection process, both undergraduate and graduate programs, to obtain a standard question that

is capable of selecting participants who have performed successfully in the study of the probability of future success.

Regarding the review of the initial knowledge of registered participants who have already as a student in UIN Sunan Kalijaga Yogyakarta, we tried to look for from GPA Result. Table 2 has shown the achievement of GPA at the beginning of the semester. The result has denoted 60.5% with a satisfactory cluster. That evidence has denoted the student had been scaffolding self-regulated learning, which is responsible for their task. This responsibility as a part of the character which keeps up Rasul trait, Fatonah, Amanah, Shidiq, Tabliq (FAST). That trait must be pursued by the student as a leader in the future. The first evidence appeared from the result of GPA in the first years which described the high achievement particularly in a score of GPA. This result showed the instrument that has been developed is the potential to find potential prospective students who have high competencies in achieving Expected Learning Outcome (ELO) in UIN Sunan Kalijaga Yogyakarta. Expected Learning Outcome (ELO) is a minimum competency, which has to be achieved by the students. This competency has a connection with a 21st-century skill, which pushes up on the skill, knowledge, expertise with the final purpose is to success in work, life, and citizenship.

D. Conclusion

In conclusion, the test instrument for enrollment in the State Islamic University has been developed, i.e., the Academic Potential Ability Test, Arabic Proficiency Test, and English Proficiency Test. Those instruments meet the validity and reliability requirements of GOOD from professional assessor and item analysis as well as has discriminating power in High level (≤ 1), the level of difficulty is distributed in proportional range, the quality of the distractor is relatively good enough, and able to measure the performance/ability of the participants in the future (predict the future success). The selection process conducted in UIN Sunan Kalijaga Yogyakarta is similar to the general university. The evidence of GPA denotes that the student had been scaffolding self-regulated learning, which is responsible for their task. The next study that can be held by the admission unit in university is improving the big storage of item tests that have exactly the proportion of distribution degree of difficulties. Consequently, if the admission will offer a repetition of selection, the items test will measure the same potential ability of the participant.

REFERENCES

Azwar, S. (2009): Dasar-dasar psikometri [The basics of psychometric]. Yogyakarta Indonesia: Pustaka Pelajar.

Brewis, Elisa (2019): Fair access to higher education and discourses of development: a policy analysis from Indonesia. In *Compare: A Journal of Comparative and International Education* 49 (3), pp. 453–470. DOI: 10.1080/03057925.2018.1425132.

Collins, Terrence J. (2017): Review of the twenty-three-year evolution of the first university course in green chemistry: teaching future leaders how to create sustainable societies. In *Journal of Cleaner Production* 140, pp. 93–110. DOI: 10.1016/j.jclepro.2015.06.136.

Derrington, Mary Lynne; Campbell, John W. (2015): Implementing new teacher evaluation systems: Principals' concerns and supervisor support. In *J Educ Change* 16 (3), pp. 305–326. DOI: 10.1007/s10833-015-9244-6.

DeVellis, Robert F. (2012): Scale development. Theory and applications. 3rd ed. Thousand Oaks, Calif.: SAGE (Applied social research methods series, 26).

Enright, Mary K.; Powers, Donald E. (1990): Validating The Gre Analytical Ability Measure Against Faculty Ratings Of Analytical Reasoning Skills. In *ETS Research Report Series* 1990 (2), i-29. DOI: 10.1002/j.2333-8504.1990.tb01358.x.

Erzad, Azizah Maulina (2016): Inovasi Dalam Pembelajaran Mata Kuliah Praktikum Bahasa Inggris Di Stain Kudus. In *EDUKASIA* 10 (2). DOI: 10.21043/edukasia.v10i2.798.

Gravetter, Frederick J.; Wallnau, Larry B. (2017): Statistics for the behavioral sciences. Edition 10. Boston, MA: Cengage Learning.

Hambleton, Ronald K.; Jones, Russell W. (1993): An NCME Instructional Module on. In *Educational Measurement: Issues and Practice* 12 (3), pp. 38–47. DOI: 10.1111/j.1745-3992.1993.tb00543.x.

Jaiswal, Garima; Sharma, Arun; Yadav, Sumit Kumar (2019): Analytical Approach for Predicting Dropouts in Higher Education. In *International Journal of Information and Communication*

Technology Education 15 (3), pp. 89–102. DOI: 10.4018/IJICTE.2019070107.

Jehanzeb, K., & Bashir, N.A (2013): Training and Development Program and its Benefits to Employee and Organization: A Conceptual Study 5 (2), pp. 243–252. Available online at https://www.iiste.org/Journals/index.php/EJBM/article/view/3947/4005.

Joe, Hye-Kyoung; Hiver, Phil; Al-Hoorie, Ali H. (2017): Classroom social climate, self-determined motivation, willingness to communicate, and achievement: A study of structural relationships in instructed second language settings. In *Learning and Individual Differences* 53, pp. 133–144. DOI: 10.1016/j.lindif.2016.11.005.

Kaplan, Robert M.; Saccuzzo, Dennis P. (2013): Psychological Testing. 8th ed. Belmont, CA: Cengage Learning US.

Keeves, John P.; Alagumalai, Sivakumar (2005): Applied Rasch measurement. A book of exemplars: papers in honor of John P. Keeves. Dordrecht, Norwell, MA: Springer (Education in the Asia-Pacific region series, v. 4).

Kern, Margaret L.; Waters, Lea E.; Adler, Alejandro; White, Mathew A. (2015): A multidimensional approach to measuring well-being in students: Application of the PERMA framework. *The journal of positive psychology* 10 (3), pp. 262–271. DOI: 10.1080/17439760.2014.936962.

Khajavy, Gholam Hassan; MacIntyre, Peter D.; Barabadi, Elyas (2018): Role Of The Emotions And Classroom Environment In Willingness To Communicate. In *Stud Second Lang Acquis* 40 (3), pp. 605–624. DOI: 10.1017/S0272263117000304.

Kline, Paul (2015): Handbook of test construction. Introduction to psychometric design: Routledge (Psychology revivals).

Kyle, V.A., White, K.M., Hyde, M.K., Occhipinti, S. (2014): The role of goal importance in predicting university students' high academic performance. In *Australian Journal of Educational & Developmental Psychology* 14, pp. 17–28. Available online at https://www.newcastle.edu.au/__data/assets/pdf_file/0003/132843/V14_2_Kyle_et_al.pdf.

Laurillard, Diana (2007): Rethinking university teaching. A conversational framework for the effective use of learning technologies. 2nd ed., reprinted. London: Routledge.

Li, Feiming; Cohen, Allan; Bottge, Brian; Templin, Jonathan (2016): A Latent Transition Analysis Model for Assessing Change in Cognitive Skills. In *Educational and psychological measurement* 76 (2), pp. 181–204. DOI: 10.1177/0013164415588946.

Mardapi, D. (2008): Teknik Penyusunan Instrumen Tes dan Nontes [The Technique of Preparing Test and Non-Test Instruments].: Mitra Cendikia Press Indonesia.

Mertens, Donna M. (2015): Research and evaluation in education and psychology. Integrating diversity with quantitative, qualitative, and mixed methods. Fourth edition. Thousand Oaks, CA: SAGE Publications, Inc.

Monteiro, Vera; Mata, Lourdes; Peixoto, Francisco (2015): Intrinsic Motivation Inventory: Psychometric Properties in the Context of First Language and Mathematics Learning. A *Psicol. Reflex. Crit.* 28 (3), pp. 434–443. DOI: 10.1590/1678-7153.201528302.

Mujab, Syaiful (2016): Evaluasi Proses Pembelajaran Reading Pada Mata Kuliah Bahasa Inggris Ii Di Jurusan Tarbiyah Stain Kudus. In *EDUKASIA* 9 (2). DOI: 10.21043/edukasia.v9i2.780.

Mulyasa, E. (2016): Improved quality management based learning for preparing the character of graduates in response to the globalization era 4 (11). Available online at https://www.ijern.com/journal/2016/November-2016/34.pdf.

Neisser, Ulric; Boodoo, Gwyneth; Bouchard, Thomas J., JR.; Boykin, A. Wade; Brody, Nathan; Ceci, Stephen J. et al. (1996): Intelligence: Knowns and unknowns. In *American Psychologist* 51 (2), pp. 77–101. DOI: 10.1037//0003-066X.51.2.77.

Nilson, Linda Burzotta (2010): Teaching at its best. A research-based resource for college instructors. Third edition. San Franciso: Wiley.

Noaman, Amin Y.; Ragab, Abdul Hamid M.; Madbouly, Ayman I.; Khedra, Ahmed M.; Fayoumi, Ayman G. (2017): Higher education quality assessment model: towards achieving educational quality standard. In *Studies in Higher Education* 42 (1), pp. 23–46. DOI: 10.1080/03075079.2015.1034262.

Privitera, Gregory J. (2016): Research methods for the behavioral sciences. [Place of publication not identified]: Sage Publications.

Sithole, Alec; Chiyaka, Edward T.; McCarthy, Peter; Mupinga, Davison M.; Bucklein, Brian K.; Kibirige, Joachim (2016): Student Attraction, Persistence, and Retention in STEM Programs: Successes and Continuing Challenges. In *HES* 7 (1), p. 46. DOI: 10.5539/hes.v7n1p46.

Slavin, R. E. (2011): Educational Psychology. Theory and Practice, Student Value Edition: Pearson College Div.

Sukarjo, & Sari, L.P. (2008): Penilaian Hasil Belajar Kimia [Assessment of Chemistry Learning Outcomes].: UNY Press Yogyakarta Indonesia.

Vickers, M., Finger, L., Barker, K., Bodkin-Andrews, G. (2014): Measuring the impact of students' social relations and values: Validation of the social-relational support for education instrument 14, pp. 71–92. Available online at https://www.newcastle.edu.au/__data/assets/pdf_file/0010/139078/self-3-vickers-et-al-2014.pdf.

Wiliam, Dylan (2011): What is the assessment for learning? In *Studies in Educational Evaluation* 37 (1), pp. 3–14. DOI: 10.1016/j.stueduc.2011.03.001.