



Biometric and Intelligent Self-Assessment of Student Progress system

A. Kaklauskas^{a,*}, E.K. Zavadskas^a, V. Pruskus^b, A. Vlasenko^a, M. Seniut^a, G. Kaklauskas^c,
A. Matuliauskaite^a, V. Gribniak^c

^a Research Institute of Internet and Intelligent Technologies, Vilnius Gediminas Technical University, Sauletekio al. 11, LT-10223 Vilnius, Lithuania

^b Department of Philosophy and Political Theory, Vilnius Gediminas Technical University, Sauletekio al. 11, LT-10223 Vilnius, Lithuania

^c Department of Bridges and Special Structures, Vilnius Gediminas Technical University, Sauletekio al. 11, LT-10223 Vilnius, Lithuania

ARTICLE INFO

Article history:

Received 12 December 2009

Received in revised form

20 March 2010

Accepted 25 March 2010

Keywords:

e-Learning

Voice stress analysis

Intelligent system

e-Self-assessment

e-Examination

Historical information

Reliability of results

ABSTRACT

All distance learning participants (students, professors, instructors, mentors, tutors and the rest) would like to know how well the students have assimilated the study materials being taught. The analysis and assessment of the knowledge students have acquired over a semester are an integral part of the independent studies process at the most advanced universities worldwide. A formal test or exam during the semester would cause needless stress for students. To resolve this problem, the authors of this article have developed a Biometric and Intelligent Self-Assessment of Student Progress (BISASP) System. The obtained research results are comparable with the results from other similar studies. This article ends with two case studies to demonstrate practical operation of the BISASP System. The first case study analyses the interdependencies between microtremors, stress and student marks. The second case study compares the marks assigned to students during the e-self-assessment, prior to the e-test and during the e-test. The dependence, determined in the second case study, between the student marks scored for the real examination and the marks based on their self-evaluation is statistically significant (the significance >0.99%). The original contribution of this article, compared to the research results published earlier, is as follows: the BISASP System developed by the authors is superior to the traditional self-assessment systems due to the use of voice stress analysis and a special algorithm, which permits a more detailed analysis of the knowledge attained by a student.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

The e-learning Master degree studies program, “Real Estate Management”, was introduced at Vilnius Gediminas Technical University in 1999, the Master degree studies program, “Construction Economics”, from 2000 and the Master degree studies program, “Internet Technologies and Real Estate Business” from 2003 (see <http://odl.vgtu.lt/>). There are currently 220 master students from all over Lithuania studying in these three e-learning master programs.

Various methodologies (Big Five Factors and Five Factor Model, intelligence quotient tests, self-assessment) are employed to predict a student’s academic motivation, achievement and academic attainments. However, in the opinions of academics and practitioners alike, more than merely a student’s personal attributes affects that student’s advancement. Rutter and Maughan (2002) hold the opinion that the quality of a school makes a significant difference on a student’s progress. A concise analysis of each of these methodologies follows.

A number of academics in the world (Komarraju, Karau, & Schmeck, 2009; Nofle & Robins, 2007) analyzed the role of the Big Five Factors (Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism) for predicting the academic motivation, achievement and academic attainment of students.

Nofle and Robins (2007) examined the relationships between the Big Five personality traits and academic attainments, specifically SAT scores and grade-point averages (GPA). Openness was the strongest predictor of SAT verbal scores, and Conscientiousness was the strongest predictor of GPAs both at the high school and college levels. These relationships replicated across four independent samples and across four different personality inventories. Further analyses showed that Conscientiousness predicted college GPA, even after controlling for

* Corresponding author. Tel.: +370 5 2745234; fax: +370 5 2745235.

E-mail address: arturas.kaklauskas@vgtu.lt (A. Kaklauskas).

secondary school GPA and SAT scores and that increased academic effort and higher levels of perceived academic ability intervened both concurrently and longitudinally in the relationship between Conscientiousness and college GPA (Noftle & Robins, 2007).

Komaraju et al. (2009) analyzed the role of the Big Five personality traits for predicting academic motivation and achievements among college students. College undergraduates (308 students) completed the Five Factor Inventory and the Academic Motivations Scale and reported their college grade point average (GPA). A correlation analysis reveals an interesting pattern of significant relationships. Further four personality traits (Conscientiousness, Openness, Neuroticism and Agreeableness) explain 14% of the variances in GPA, and an intrinsic motivation for accomplishment explained 5% of the variances in GPA. Finally Conscientiousness emerges as a partial mediator of the relationship between intrinsic motivation to accomplish and GPA. These results are interpreted in the context of what educators can do to encourage and nurture student motivation and achievement (Komaraju et al., 2009).

Although IQ scores are good predictors of academic achievement in elementary and secondary school, the correspondence between IQ and academic performance is less consistent at higher levels of education, and many have questioned the ability of IQ tests to predict success later in life. The tests don't measure many of the qualities necessary for achievement in the world of work, such as persistence, self-confidence, motivation, and interpersonal skills, or the ability to set priorities and to allocate one's time and effort efficiently. In addition, the creativity and intuition responsible for great achievements in both science and the arts are not reflected by IQ tests (Intelligence Quotient, 2009). As education progresses, other factors (motivation, opportunity, organisation, background, intelligence, teaching) come into play, such that the limits are determined by the strength of the weakest link in the chain (Intelligence, 2009).

Self-assessment in an educational setting involves students making judgments about their own work. Students can make assessment decisions regarding their own essays, reports, projects, presentations, performances, dissertations and even exams. Self-assessment can be extremely valuable in helping students critique their own work and form judgments about its strengths and weaknesses. For obvious reasons, self-assessment is usually employed as part of a formative assessment rather than a summative process which requires certification by others (Self-assessment, 2009). Academics and practitioners (Austin, Gregory, & Galli, 2008; Ballantine, McCourt Larres, & Oyeler, 2007; Sundström, 2008; Sung, Chang, Chang, & Yu, 2009 & Zimmerman & Kitsantas, 2007) performed tests on self-assessment in an educational setting with groups of various ages (teenagers, college students, adults and such) in various areas (schools, universities, workplaces).

Many academics are seeking to diversify assessment tasks, broaden the range of skills assessed and provide students with more timely and informative feedback on their progress. Others wish to meet student expectations for more flexible delivery and to generate efficiencies in assessment that can ease academic staff workloads. As more students seek flexibility in their courses, it seems inevitable that there will be growing expectations for flexible assessment as well (Formative Assessment, 2009). The said reasons and the desire to protect students against psychological traumas during their progress assessment have encouraged us to develop the Biometric and Intelligent Self-Assessment of Student Progress (BISASP) System.

The Biometric and Intelligent Self-Assessment of Student Progress system is developed on the basis of the aforementioned methodologies (Big Five Factors and Five Factor Model, intelligence quotient tests, self-assessment), the thirteen years of experience with distance learning by the authors of this article and the technology for biometric voice stress analysis. These methodologies are very helpful and significant, while the results of their practical application are often, though not always, reliable. Therefore, the analysis how the Big Five Factors, intelligence quotient tests and self-assessment helped to deal with the reliability issues of self-assessment in practical applications was particularly significant in development of the BISASP System. Also special attention was paid to self-, peer- and teacher-assessments (the criteria systems used, integration of such criteria into one general assessment, use of aggregation methods, the level of reliability of the results, tendencies in the results gained) and to a comparison of this assessment with actual learning results.

The structure of this paper is as follows. Following this introduction, Section 2 describes self-assessment. Section 3 provides a description of the Biometric and Intelligent Self-Assessment of Student Progress System. Sections 4 and 5 provide a brief review of the Self-assessment Integrated Scoring Model and the Self-assessment Integrated Scoring Adjustment Model. Section 6 follows with case studies. Finally the concluding remarks appear in Section 7.

2. Reliability of self-assessment

Numerous studies have been conducted worldwide which analyze the reliability of self-assessment. There are many controversial opinions on this issue. For example, in the opinion of Matsuno (2009), a number of researchers have reported high correlations between student- and teacher-assessments, while other studies have shown low correlations between them. Despite meta-analyses of self-assessment in higher education deeming that students will be able to self-assess accurately, the reported correlations between self and tutor evaluations within medical PBL programs are uniformly low (Papinczak, Young, Groves, & Haynes, 2007). A great many researchers (AlFallay, 2004; van Braak, 2004; Fitzgerald, White, & Gruppen, 2003; Marsh, Overall, & Kesler, 1979; Mynttinen et al., 2009; Sung, Chang, Chiou, & Hou, 2005; Xiao & Lucking, 2008 and others) reached reliable results proving that the reliability of self-assessment is sufficient. Several examples of such research are further presented.

Xiao and Lucking (2008) examined the validity and reliability of student generated assessment scores. Two hundred and thirty two, predominantly undergraduate students were selected by convenience sampling during the fall semester of 2007. The findings indicate that the validity and reliability of student generated rating scores were extremely high.

AlFallay (2004) investigates the role of some selected psychological and personality traits of learners of English as a foreign language for accuracy in their self- and peer-assessments. AlFallay (2004) obtained high reliability indices similar to those reported in the literature. The study also shows that long periods of practice and sufficient feedback have a positive effect on the accuracy of assessments. Finally the study demonstrates that students with low self-esteem are the most accurate in assessing their performance, whereas learners with instrumental motivation are the least accurate (AlFallay, 2004).

Although studies have examined the abilities of medical students to self-assess their performance, there are few longitudinal studies that document the stability of self-assessment accuracy over time (Fitzgerald et al., 2003).

Ballantine et al. (2007) examined the reliability of self-assessment as an indicator of students' knowledge across a variety of disciplines. As per the opinions of Ballantine et al. (2007), an agreement between self-evaluation and alternative measures of assessment has been reported in a number of studies.

Sung et al. (2005) describe the web-based self- and peer-assessments system. Sung et al. (2005) show that significant consistency is found between the results of student self- and peer-assessments and the results of teacher assessments.

Marsh et al. (1979) analyzed a comparison of faculty self-evaluations and the evaluations by their students. Considerable student–faculty agreement was found when 83 college courses were evaluated. Separate factor analyses indicated similar dimensions in both student and faculty evaluations. Validity coefficients were significant for all evaluation factors. Mean differences between student and faculty ratings were low, and the two groups agreed on the behaviors most descriptive of the faculty (Marsh et al., 1979).

Mynttinen et al. (2009) examined overconfidence among novice drivers by comparing their self-assessed driver competence with the assessments made by driving examiners. A Finnish ($n = 2739$) and a Dutch sample ($n = 239$) of driver's license candidates assessed their driver competence in six areas and then took the driving test. In contrast to previous studies where drivers assessed their skills in comparison to an average driver, a smaller proportion overestimated and a larger proportion made realistic self-assessments of their driver competence in the present study, where self-assessments were compared with examiner assessments. Between 40% and 50% of the candidates in both samples made realistic assessments, and 30–40% overestimated their competence.

Ballantine et al. (2007) evaluate the reliability of self-assessment as a measure of computer competence. The scores achieved by students in self-assessed computer competence tests are compared with scores achieved in objective tests to evaluate the reliability of self-assessed computer competence. The results reveal a statistically significant over-estimation of computer competence among the surveyed students.

Braak (2004) analyzed the self-perceived computer competence of university students. Questionnaires were administered to two samples of first-year university students in psychology and education. The results indicated a high correlation between the two scales of computer competence (Braak, 2004).

The System developed by the authors of this article is also based on the same presumption that, by assigning students questions in accordance with some certain methodology (entirely unrelated with the contents of the exam) and then processing them in accordance with a certain algorithm, it is possible to determine a student's level of knowledge rather accurately. In other words, internally a student senses the rating that his/her knowledge is worth. The existing experience, intuition and analysis of the learning process give students a rather accurate indication of what mark they can expect to earn on an exam.

What makes the System developed by these authors superior to the traditional systems is the use of biometric voice stress analysis which permits a more detailed analysis of the knowledge a student has assimilated.

3. Biometric and Intelligent Self-Assessment of Student Progress (BISASP) System

The human voice reflects that person's spiritual and emotional state of being, motivation and preparedness to complete an assignment. The internal resolve of an individual to complete an assignment (an assessment of the self and one's chances for doing so) and an external evaluation according to corresponding objective criteria always differ. This is why various tensions arise between an examiner and the person being examined during an exam. These tensions are caused by the different imaginings regarding the degree of preparedness for the exam; the self-assessment of an individual taking the exam influences such an imagining. In an effort to decrease such tension, restore an educational sphere and teach collegial team work, tolerance and the goal of a better and more objective appraisal of knowledge and abilities, the authors here developed the Biometric and Intelligent Self-Assessment of Student Progress (hereafter – BISASP) System.

The test questionnaire is composed to contain three dimensions of a student's preparedness to take an exam—the psychological assessment of one's own preparedness to take an exam, the objective assessment of preparedness and expectations.

Our research was based on two hypotheses: (1) a well-prepared self-assessment process may enable a rather reliable forecasting of student academic achievement; (2) microtremor frequencies depend on the student's stress and, therefore, of the level of his/her readiness for the exam. Questionnaires were used as the method of data collection. Each question in the questionnaires was formulated and their evaluation procedure was set enabling students to assess themselves on a 10-point scale during the self-assessment. It facilitated estimation of their self-rated forecasted exam marks. Application of a specific corresponding algorithm establishes how much a student's self-assessment and expectations conform to the realistic level of that student's preparedness.

Furthermore, our research revealed that such self-assessment done during the psychological test helps to adjust a module to student needs better.

Initially students answer thirteen questions both verbally and electronically; these questions are not directly related with the learning contents. Afterwards the students complete an electronic questionnaire which serves as the basis for establishing the level of a student's self-assessment (high, average, low). By employing a special algorithm, the system evaluates the microtremors and digital information submitted by a student and writes a subjective rating of a psychological assessment. After this the students are able to take a real exam electronically and compare the received subjective and objective rating marks amongst themselves. This model can be employed for evaluating student progress and, additionally, for resolving conflictual situations.

The authors of this article have developed a voice stress database which contains the answers given by students during an e-psychological test and a specific algorithm, the core of the BPASP System, which can evaluate a student's knowledge by assigning a subjective rating mark to the psychological test performed prior to the exam.

The Biometric and Intelligent Self-Assessment of Student Progress (BISASP) System consists of the following components (see Fig. 1): Database Management Subsystem and Databases, Equipment Subsystem, Model-base Management Subsystem and the Model Bases with User Interface. The components of BISASP System are briefly analyzed below.

Sound recording and data input equipment: microphone, sound card and a PC for recording audio files. Sound recorder software: mostly software that converts sound recordings of different formats and media and saves such recordings as 16 bit 44.1 kHz WAV files on HDD. Sound processing equipment: either software or hardware for audio file/signal processing in order to single out the parameters that help to determine emotional state. Such processing may be done either by hardware with special controllers for Fourier transforms, wavelet transforms, determination of MFCC coefficients and other operations or by software using special applications for digital signal processing.

The Database contains the developed Historical statistics database, Domain database, Question database, Self-assessment DB, Self-assessment Integrated Rating DB, Self-assessment Adjusted Integrated Rating DB, Examination results DB and Computer learning systems database.

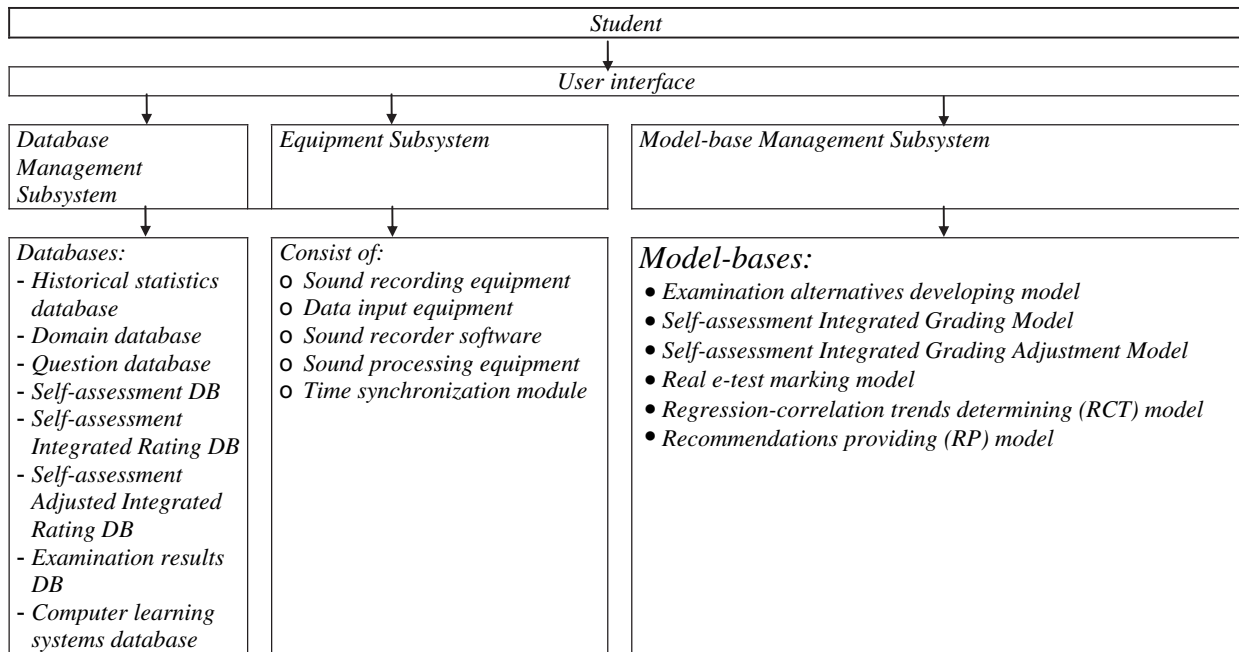


Fig. 1. Composite parts of the Biometric and Intelligent Self-Assessment of Student Progress (BISASP) system.

The *Historical statistics database* accumulates historical statistical data gathered from the Self-assessment Integrated Grading Model, Self-assessment Integrated Grading Adjustment Model and real e-test (testing questions, testing results [information on correct and incorrect answer, time distribution to every question, number of times a student has changed an answer to each question of a test], voice stress data, correlation between emotional stress, correct answers data and such).

The *Domain database* contains information and knowledge that an instructor is teaching. Over three semesters in the master study courses, students complete seven core modules and five optional modules. Students choose an elective from 21 modules within “Real Estate Management” and 17 modules within “Construction Economics”, both Master’s degree study programs, and they should optionally pass five examinations. During the fourth semester, master students write a final thesis. After registration, students mark the sections of the elective modules they want to study onto electronic questionnaires. The *Domain database* also offers study materials to students according to the repetitive key words in the different optional modules. A mixed approach is also possible and available. The received information is used for action plans, i.e., “mini curricula” that are used to guide the learner/student.

The *Question database* accumulates the following information: questions according to modules, possible answers to a question and evaluation of the correctness of possible answer versions.

The *Computer learning systems database* enables use of different Web-based computer learning systems: construction, real estate, facilities management, international trade, ethics, innovation, sustainable development, building refurbishment and others (Kaklauskas, Ditkevicius, & Gargasaitė, 2006; Kaklauskas, Kaklauskas, Krutinis, & Ditkevicius, 2005; Kaklauskas, Zavadskas, & Ditkevicius, 2006; Kaklauskas et al., 2007; Zavadskas & Kaklauskas, 2000; Zavadskas, Kaklauskas, Vainiunas, & Gikys, 2000; Zavadskas, Kaklauskas, & Vlasenko, 2008; Vainiunas & Kaklauskas, 2005).

The *Equipment Subsystem* consists of sound recording equipment, data input equipment, sound recorder software, sound processing equipment and a time synchronization module.

The *Model-base* consists of six models: the Examination alternatives developing model, Self-assessment Integrated Grading Model, Self-assessment Integrated Grading Adjustment Model, Real e-test marking model, Regression–correlation trends determining model and Recommendations providing model.

The *Real e-test marking model* provides several possible answers for students to select as the best for the multiple-choice questions during the time of the real testing. An incorrect answer is denoted as 0 and a correct answer, as 1; intermediate answers are scored from 0 to 1. The difficulty of a question is determined on the basis of the results from previous tests taken by other students.

The *Regression–correlation trends determining (RCT) model* was developed by these authors as an aid to collect and organize numerical information into tables, graphs and charts, to analyze and interpret numerical data and to make informed decisions. By applying the RCT Model, it is possible to discover interesting patterns that identify a student’s behavior on the system and to store student interactions and feedbacks in the BISASP System; thereby past memory experiences are maintained, and then new teaching paths are derived. The RCT Model provides information on a testing process in matrix and graphical form including information on correct and incorrect answers, time distribution for every question, number of times a student changes an answer to a test question and the like. Complex parameters are also presented whereby not only is the correctness of an answer evaluated but also the time required for a student to answer it along with hesitations of selection. The knowledge assessment can possibly change once an answer is evaluated by a complex parameter. The RCT Model, which is based on statistical information accumulated from the Self-assessment Integrated Grading Model, Self-assessment Integrated Grading Adjustment Model and the real e-test, presents various regression–correlation dependencies between different parameters and data. Some such dependencies are presented and explained in the Case Study.

By having such integrated information on the testing process along with the Question database, it is possible to create tests in a nonrandom manner that individualizes tests for each student according to the number of questions, their difficulty and the proportion of questions from different subjects. This is performed by the *Examination alternatives developing model*. The gained test results are stored in the Examination results database. Students can use the statistics provided by the RCT Model to see the difficulty of a question and the average assessment for the entire group; thus students are able to learn about their position in the group before and after their studies. Once data is stored about the difficulty of questions, an opportunity opens to offer easier questions first and later to move onto more difficult ones. The question subjects can also be selected similarly, by moving from a simpler subject to a more difficult one and then repeating the most difficult. Therefore, by forming a compiled base of questions, the questions for tests are not formulated randomly but individually adapted for each student according to the number of questions, their difficulty and the proportion of questions from different modules. Furthermore easier questions can be given at the beginning, and then the test can proceed to the more difficult questions. Similarly the subjects taught can be selected from the easier to the more complex, and subjects that are not yet mastered can be repeated.

The *Recommendations providing (RP) model* collects information on the history of a student's responses, provides feedback which helps to determine the strengths and weaknesses of that student's knowledge and then provides various recommendations for further education. The RP Model explains why one or another answer is incorrect and offers use of certain additional literature and multimedia to clarify the incorrectly answered questions. Analogically the RP Model, since its basis is integrated information on the testing process, can show areas of improvement to the instructor of a module. If, for example, more than 200 students spend, on average, more than 25% of their time answering the test questions in the "Real Estate Market Analysis" section (as compared to the rest of the modules), and the rating marks for this section of the exam are more than 2 points lower than the average number of points for the module, then the RP Model makes recommendations to the instructor to supplement and more thoroughly explain the more difficult areas and the like.

The system provides a student-friendly *interface* to facilitate use of the teaching services easily. It provides an introduction to the system, a comprehensive window based menu of services and other relevant information.

Psychological examination consists of two modules:

- Self-assessment Integrated Grading Model.
- Self-assessment Integrated Grading Adjustment Model.

Next, a more detailed analysis of how the integrated self-assessment mark is estimated during a student's self-assessment is provided as an example.

4. Self-assessment Integrated Grading Model

Self-assessment Integrated Grading Model was being developed, similar studies that were conducted worldwide were analyzed (Big Five Factors and Five Factor Model, intelligence quotient tests, self-assessment and others). Special attention was paid to self-, peer- and teacher-assessments (the criteria systems used, integration of such criteria into one general assessment, use of aggregation methods, the level of reliability of the results, tendencies in the results gained) and to a comparison of this assessment with actual learning results.

By basing their work on the aforementioned and other studies, the authors of this work performed the following:

- Formulated a 16-questions/criteria system to use as the basis to conduct the psychological tests,
- Utilized the Voice Stress Analyzer,
- Developed multi-criteria analysis methods as the basis for integrating the criteria under discussion into the psychological test assessment,
- Developed the model for assigning a rating mark to the e-psychological test on the basis of the formulated criteria system, the developed methods and the Voice Stress analyzer.

The analysis of questions used in self-evaluation which were developed by researchers (Komarraju et al., 2009; Noftle & Robins, 2007) from various countries helped the author to create 16-questions/criteria system of their own. For example, Noftle and Robins (2007) examined the relationships between the Big Five personality traits and academic attainments, specifically SAT scores and grade-point averages (GPA). Conscientiousness was the strongest predictor of GPAs (Noftle & Robins, 2007). Komarraju et al. (2009) analyze the role of the Big Five personality traits for predicting academic motivation and achievements among college students. These results are interpreted in the context of what educators can do to encourage and nurture student motivation and achievement (Komarraju et al., 2009). As education progresses, other factors (motivation, opportunity, organisation, background, intelligence, teaching) come into play, such that the limits are determined by the strength of the weakest link in the chain (Intelligence, 2009).

Next 16 criteria used as the basis for the e-self-assessment are named:

- What mark are you expecting for the exam?
- How many days have you spent studying?
- What was your average for the last semester?
- In your opinion, is the subject you've been studying essential and necessary for your future profession?
- Were the studies of this course interesting?
- Did the time you spent studying this discipline correspond with the benefit you received?
- What average do you expect for this semester?
- What was your average for your undergraduate studies?
- What percent of the exam material have you assimilated?
- Did you attend all lectures and practice sessions?

- How many days did you need to prepare for this exam to get the expected mark?
- How many days, in average, did you need to prepare for past exams to get the expected mark?
- What is your accuracy, as a percentage, at guessing your exam mark?
- If this subject was an elective, and you were able to choose, would you have chosen this or some other subject?
- Did the instructor present the material understandably and clearly?
- Was the subject you studied personally interesting and important to you?

It was noticed that, during the time of the self-assessment, students experimented with the BISASP System. They would submit answers that had nothing to do with reality, ones they'd dreamed up. Such falsified answers were eliminated using the Voice Stress Analyzer. Whenever microtremor frequencies that exceeded the limit of 11 Hz were discovered in a student's voice, no attention was paid to that answer.

The authors here developed multi-criteria analysis methods as the basis for integrating the criteria under discussion into the sum of the psychological test assessment as follows:

1. A new method for the complex determination of the weight of the criteria was developed taking into account their quantitative and qualitative characteristics. This method allows calculating and coordinating the weights of the quantitative and qualitative criteria according to the aforementioned characteristics.
2. A new method of a multiple criteria complex, proportional evaluation of results from the psychological test was suggested that would enable an examiner to obtain a reduced criterion to determine the complex (overall) efficiency of the test. This generalized criterion is directly proportional to the relative effect of the values and weights of all the criteria considered.
3. The results from the psychological test were reviewed in consideration of the psychological test results of other students whose difference between the results from their psychological and their real test was not greater than 0.5 points. According to this method, the result from the psychological test being estimated is directly proportional to the system of adequately descriptive criteria and the values and weights of such criteria.

Self-assessment Integrated Grading Model was developed on the basis of the formulated criteria system, developed multiple criteria analysis methods and the Voice Stress Analyzer.

5. Self-assessment Integrated Grading Adjustment Model

Undoubtedly other individuals and socio-cultural factors influence self-assessment. These include the temperament of an individual, social conditions of life, features of interactions and communications in society, support for taking initiatives and others (Trandis, 1989). Despite this, many researchers are inclined to distinguish the major levels of self-assessment which are applicable to members of all socio-cultural communities. These are the *high, average and low levels of self-assessment*. Actually certain authors indicate another two levels—*too high and too low*.

Each one of the aforementioned types of self-assessment has characteristics of its own.

5.1. Characteristics of high self-assessment

- Overconfidence in one's own strengths
- Desire to change one's social status and situation
- Inclination to make risky decisions
- Exploitation of all opportunities in the pursuit of a defined goal
- Desire to explain and look for the "guilty" when facing a lack of success

5.2. Characteristics of average self-assessment

- Critical (self-critical) outlook on the self and surrounding others
- Raising realistic goals
- Ability to coordinate one's own possibilities with the demands raised by life
- Rejection of unmeasured risk
- Search for reasons in one's own actions when facing a lack of success

5.3. Characteristics of low self-assessment

- Lack of desire to make an essential change
- Learned powerlessness (let others seek a career; I'm happy with the way things are)
- Lack of confidence in one's own strengths
- Lack of desire to make independent decisions leaving that right to others

Research shows that a student's behavior and carriage during an exam, as well as expectations, depend on self-assessment. Therefore it is important to establish to what degree the student's attained knowledge and practical skills are suitable in regards to that student's real expectations.

First, the Self-assessment Integrated Scoring Adjustment Model gives 14 questions to a student and determines the student's level of self-assessment based on his/her answers. The level of self-assessment (high, average and low) is determined as a percent between 0% (the lowest self-assessment level) and 100% (the highest self-assessment level). The obtained results are used to adjust the calculations of the self-assessment integrated mark, and such mark is displayed to the student before the actual exam.

6. Case study

The e-learning, Master degree studies program, "Real Estate Management", was introduced at Vilnius Gediminas Technical University by its Department of Construction Economics and Property Management in August of 1999. The Master degree studies program, "Construction Economics", has been in effect since 2000, the Master degree studies program, "Internet Technologies and Real Estate Business", since 2003 and the Master degree studies program, "Intelligent Built Environment", since 2009 (see <http://odl.vgtu.lt/>). There are currently 239 master students from all over Lithuania studying in these four e-learning master's level programs.

Different multimedia and communication means are used during these studies, namely electronic format textbooks, video and audio equipment, computer software, computer learning systems, intelligent testing systems, intelligent tutoring system, computer conferencing, computer networks, a discussion forum and "face-to-face" contacts. To increase the efficiency and quality of e-learning studies, the Biometric and Intelligent Self-Assessment of Student Progress (BISASP) System was developed. The practical application for the previously mentioned distance learning studies is described in brief.

6.1. Case study 1: analysis on the interdependencies between microtremors, stress and student marks

According to Hans Selye's Stress Model (Selye, 1951), known as the General Adaptation Syndrome (GAS), stress develops in three stages (anxiety, resistance and exhaustion). Lippold, Redfearn, and Vuco (1957) begin exploring the correlation between muscle activity and stress. Lippold (1970) first discovers the physiological tremor in the human voice in the 8–12 Hz range.

Many studies have been performed globally that analyze the interdependency of microtremors and stress. Different scientists and practitioners do not agree with one another about this issue. Such opposing opinions are further presented in brief. Below are studies validating the technical and operational efficacy of voice stress analysis (VSA).

Under conditions of normal speech, the vocal-wave signal is modulated by a speaker in values of approximately 8 and 14 Hz when speaking in the range of 8–10 Hz. These changes in the frequency spectrum are known as microtremors. However, when a speaker is under psychological or physical pressure, the frequency of the modulation increases typically peaking in a higher range, between 10 and 12 Hz (Clarke, 2005). Ruiz, Legros, and Guell (1990) examine voice analysis to predict the psychological or physical state of a speaker.

Hopkins, Ratley, Benincasa, and Grieco (2005) have found that VSA technology can identify stress better than chance with a performance approaching that of current polygraph systems. Haddad, Walter, Ratley, and Smith (2002) determined that voice stress analysis achieved an accuracy rate of 100% when used to detect stress in 45 known-conclusion responses. Cestaro (1995) reports that the lab simulations established that the CVSA performs electrically according to the manufacturer's theory of operation and, even in the absence of jeopardy, which is a basic requirement in detection of deception.

Tippett (1994) examined 54 individuals who were convicted sex offenders on probation and in treatment for their crimes. Tippett (1994) notified that with these 54 examinations, there was a 100% agreement between the CVSA and the polygraph. The number of examinees that were found to be deceptive (DI) were 35 and the number of examinees found to be not deceptive (NDI) were 19. As a result of this study, it appears that the CVSA is as effective as polygraph, which is the question this study set out to answer (Tippett, 1994).

Ruiz et al. (1990) reports that their research suggests that psychological stress may be detected as acoustic modifications in the fundamental frequency of a speaker's voice and that the fundamental frequency of the vocal signal is slowly modulated (8–14 Hz) during speech in an emotionally neutral situation. In situations demanding increased 'mental or psychomotor' activity, the 8–14 Hz modulation then decreases as the striated muscles surrounding the vocal cords contract in response to the arousal, thus limiting the natural trembling (Ruiz et al., 1990).

Chapman (1989) selected 211 criminal responses at random from 2109 known-conclusion responses where voice stress analysis was used to test suspects. Chapman's (1989) study confirmed that voice stress analysis was accurate when utilized as a truth verification device and produced a confirmed confession rate of 96.4% where deception was indicated by VSA (Chapman, 1989).

Brockway's (1979) study reports that voice stress analysis does depict predictable and self-reported anxiety. Smith (1977) concludes that voice stress analysis is a valid measure of anxiety. Borgen and Goodman (1973) indicated that voice stress analysis of the verbal responses correlated well with the other physiological responses to acute stress.

Inbar and Eden (1976) were able to independently verify the existence of the 8–14 Hz 'microtremor' and to trace its origins to the central nervous system. Wiggins, McCranie, and Bailey (1975) concluded that audio stress can be detected with a voice stress analyzer in psychiatric patients during the course of therapy and that the VSA could serve as a useful tool for this purpose.

Heisse (1974) analyzed 91 known-conclusion criminal responses utilizing voice stress analysis and determined that audio stress analysis seems to be valid in detecting changes in various psycho physiological parameters so that a trained examiner utilizing standard techniques can evaluate these changes and thus utilize the instrument in truth and deception. Brenner (1974) utilizing a voice stress analyzer, established that vocal stress increased as a function of audience size.

It should be noted that not all researchers support this hypothesis. Bell, Ford, and McQuiston (1976) developed a device and described the method for microtremor registration. Moreover he claimed that the state of a microtremor at 8–12 Hz occurs in the human voice in a quiet, relaxed state and disappears in the case of stress.

Cestaro (1995) used the CVSA system in his research and concluded that the CVSA instrument has been shown to detect discrete changes in the fundamental frequency of speech when laboratory instruments have been used to simulate a voice microtremor. Thus Cestaro (1995) confirms NITV's underlying theory of operation. However, these results do not confirm: (1) the existence of a voice microtremor, (2) a relationship between the microtremor amplitude and psychological or physical levels of stress, (3) a reduction in microtremor amplitude during an act of deception and (4) that a voice microtremor—if it exists at all—has sufficient signal value to be detected by the CVSA (Cestaro, 1995).

In our research, the X13-VSA software was used to collect data on microtremors in voices of tested students.

The research involved a group of some 200 volunteer students who took a real, computer-based examination and had to select (tick) the correct answer on a computer display; in addition they were asked to say the selected answer aloud. During the exam, students had to mark and say the correct answers aloud to 20 questions within 10 min. The oral answers were analyzed using the VSA application, which is part of the Equipment Subsystem. During the research, a total of 4000 voice recordings in four student groups were examined and analyzed.

Based on this analysis, the BISASP System created a database with the following fields: student ID, question ID, a student's answer (correct/incorrect) and the microtremor frequency in the answer's utterances with intonational significance. This information was used to calculate the correct answers by percent and by average microtremor frequency.

The percentage of correct answers to each question was calculated using the equation, $ca = (Nc/Ta) \times 100$. Nc is the number of correct answers to a specific question. Ta is the total amount of answers (correct and incorrect) to a specific question. For example, if a certain question was answered correctly 90 times and incorrectly 60 times, then $Nc = 90$, $Ta = 90 + 60 = 150$ and $ca = (90/150) \times 100\% = 60\%$. Correct answers by percent constitute the linear trend of the Correct Answers Index for all questions asked at the experiment.

Average microtremor frequency (amtf) is the index calculated for each question using the following equation $amtf = (\sum mtf)/Ta$ (here $\sum mtf$ is the sum total of average microtremor values for all answers to a particular question, Ta is the total number of answers). For example, if a certain question was asked 7 times during the testing and the average microtremor values of significant utterances for each answer were 10.5, 11.2, 11.5, 9.8, 10.7, 11.8 and 9.5, then $amtf = (10.5 + 11.2 + 11.5 + 9.8 + 10.7 + 11.8 + 9.5)/7 = 10.7$ Hz. This index was used to arrive at a conclusion on the intensity of stress among students who answered this question: if the majority of students are stressed when answering this question (voice microtremor frequency of 10–12 Hz), then the average microtremor frequency is also high (10–12 Hz). Linear (average microtremor frequency) is the linear trend of the average microtremor frequency index for all questions asked at the experiment.

The calculated parameters were charted as presented in Fig. 2.

Fig. 2 shows the relation between correct answers and the ID numbers of the questions (bright line of the trend) and the average microtremor frequency of the answers to test questions and ID numbers of the questions (dark line of the trend). The x -axis shows the numbers of the test questions for students who had passed the exam. The left side of the y -axis shows the correct answers by percent. The right side of the y -axis shows the average microtremor frequency to each question in the exam.

In addition Fig. 2 shows two correlating curves obtained during the research:

- The direct relationship between a correct answer and the ID number of the question (broken line);
- The direct relationship between the average microtremor frequency and the ID number of the question (dark columns).

The reliability of the results was assessed by performing a correlation analysis between emotional stress (microtremor frequency) and the evaluations of correct answers (by percent) to the test questions. The analysis showed that:

- A correlation exists between emotional stress and the correctness of an answer.
- High values of microtremor frequency (11–12 Hz) were established when analyzing the oral answers to questions with a low value of correct answers. Therefore higher than average emotional stress was experienced when answering the “unknown/difficult” questions, and zero or minor emotional stress was found in the case of the “known/simple” questions.

Fig. 3 shows the results of the dependence between the examination mark and the average microtremor frequency in answers of each student.

Fig. 3 shows the relation between real mark given students during exam and students ID number (dark line of the trend) and the average microtremor frequency of the answers to test questions and students ID number (bright line of the trend). The x -axis shows the students ID number who had passed the exam. The right side of the y -axis shows the correct answers (by %). The left side of the y -axis shows the average microtremor frequency of each student.

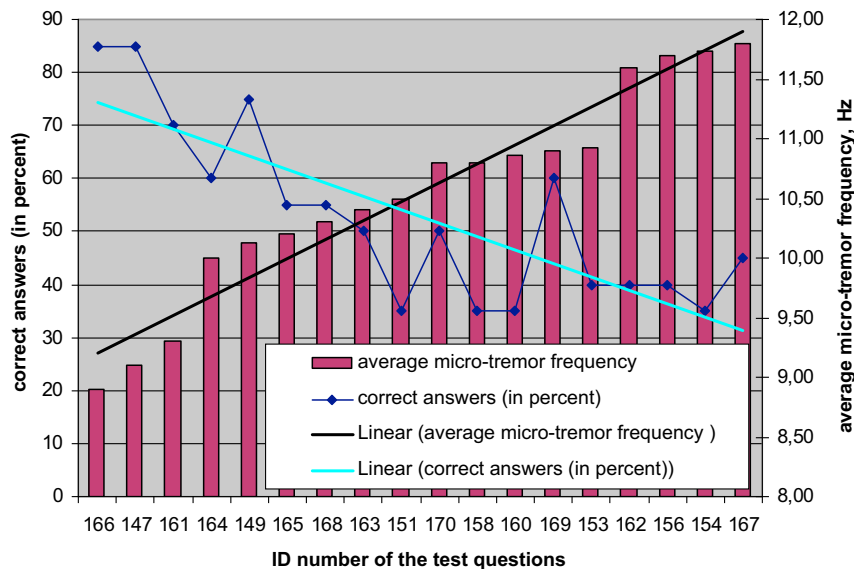


Fig. 2. Correlation between emotional stress and correct answers (by percent) in the e-test: the x -axis indicates ID numbers of the test questions and the y -axis indicates the average microtremor frequency (Hz) in a student's voice (on the right) and correct answers by percent (on the left).

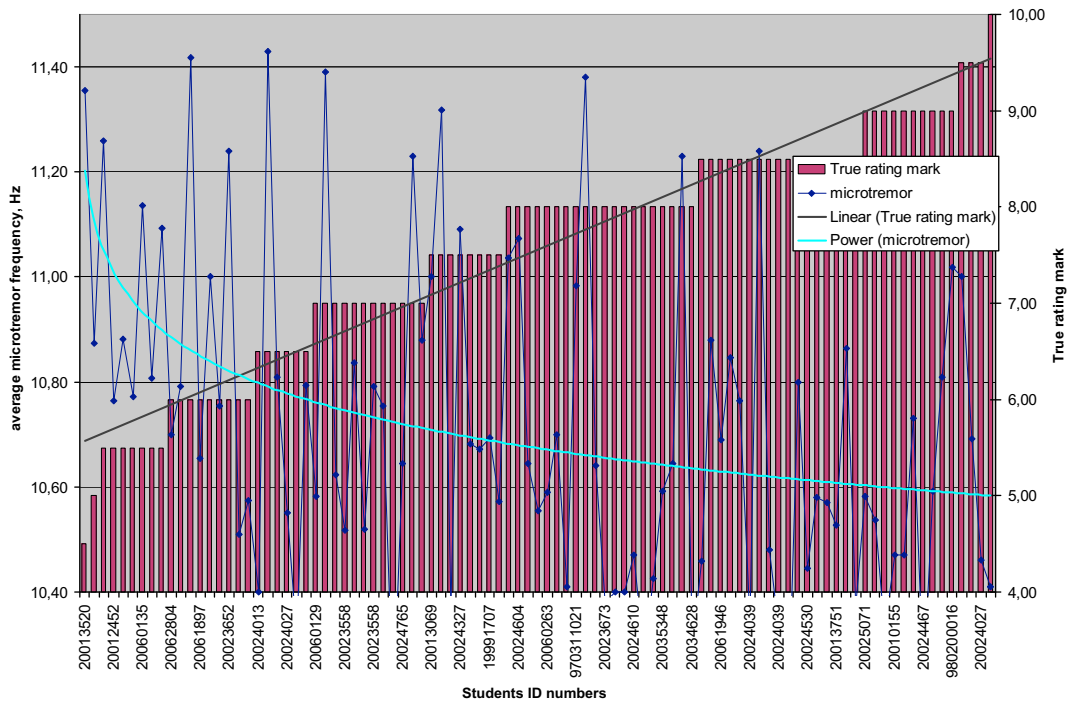


Fig. 3. Correlation between emotional stress and examination mark in the e-test: x-axis – student IDs; y-axis – examination marks of students on a 10-point scale (on the right) and the average microtremor frequency (Hz) in a student's voice (on the left).

In addition Fig. 3 shows two correlating curves obtained during the research:

- The direct relationship between exam marks and students ID numbers (dark columns);
- The direct relationship between average microtremor frequency and ID number questions (dark broken line).

The charts shown in Fig. 3 lead to the following conclusions:

- A correlation between the level of emotional stress during the examination and the marks scored in the actual examination was determined for students who took part in the experiment.
- Considerably higher stress levels (microtremor frequencies) were recorded for students who scored less than 6 points in the examination.

6.2. Case study 2: comparison of marks assigned to students during the psychological examination, prior to the e-test and during the e-test

Psychological examination consists from two modules:

- Self-assessment Integrated Grading Model.
- Self-assessment Integrated Grading Adjustment Model.

Currently students' knowledge can be automatically assessed (instead of by an exam) by using the BPASP System on the basis of student psychological tests, accumulated historical voice stress data, determined regression equation and the specially-developed algorithm. The BPASP System automatically assesses a student's knowledge before an exam according to the student's verbal/oral answers and a filled-out questionnaire. The rating mark assigned during the psychological test is also verified on the basis of the answers a student provides in the electronic, self-assessment questionnaire. For example, when a teacher/lecturer gives a student questions such as "How many days did you study?" or "Was this course interesting to learn?" or "Was the material provided by the instructor understandable and clear?" and the like before an exam, the student can be assessed precisely by using the BPASP System to assign him/her a rating mark. Fig. 4 illustrates the comparison of marks assigned to students during the e-psychological test performed prior to the exam and of the marks assigned during the exam itself (using the BPASP System). The regression–correlation curves seen in Fig. 4 show the interrelation between the marks assigned during the e-psychological test and the marks assigned during the e-examination itself.

It was established during the testing that students who had received higher rating points on an exam had poorer psychological test results than did those students who had received the higher rating points. As such a conclusion can be drawn that the students who receive high rating grades evaluate themselves as average on the psychological test. Meanwhile the students who receive the lower rating grades often over-evaluate their abilities. This is quite clearly seen in Fig. 4. Such dependencies were also noticed by Sung et al. (2009), Papinczak et al. (2007), Lejk and Wyvill (2001) and other researchers. In the opinion of Sung et al. (2009), low- and high-achieving students tended respectively to over- and underestimate the quality of their own work in self-assessments. Lejk and Wyvill (2001) present the results of an experiment which compared secret and agreed peer- and self-assessment of contributions to a group project. There was a tendency for

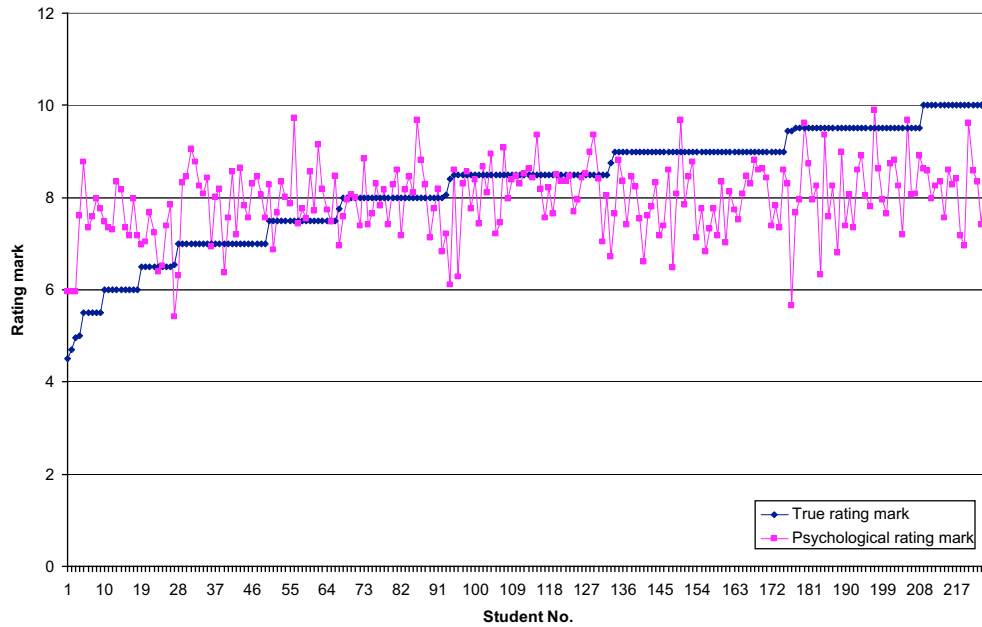


Fig. 4. Comparison of marks assigned to students during the e-psychological test, prior to the e-test and during the e-test; Legend: y-axis – rating marks of students on a ten-point scale; x-axis – student IDs; “rating marks of the e-psychological test” – marks assigned to students during the e-psychological test, prior to the e-test using the BPASP System; “marks of the real test” – actual marks assigned to students during the e-test using the Intelligent Testing System; “linear (marks of the psychological test)” – regression–correlation linear trend which describes the marks assigned to students during the e-psychological test, prior to the e-test using the BPASP System; “linear (marks of the real test)” – regression–correlation linear trend which describes the actual marks assigned to students during the real e-test.

students towards the top end of the group to under-assess themselves in comparison with the assessment by their peers and for students towards the bottom end of the group to over-assess themselves. In the opinion of Papinczak et al. (2007) opinion, the tendency exists for students to overestimate their competence, especially lower-performing students and young or highly capable students are more likely to undermark their work.

Fig. 6 presents the curves relating to the conformity of the psychological test taken by both sexes to the real test. The x-axis shows how many of the psychological test results were in full conformance with the real test results taken by female students (a) and male students (b).

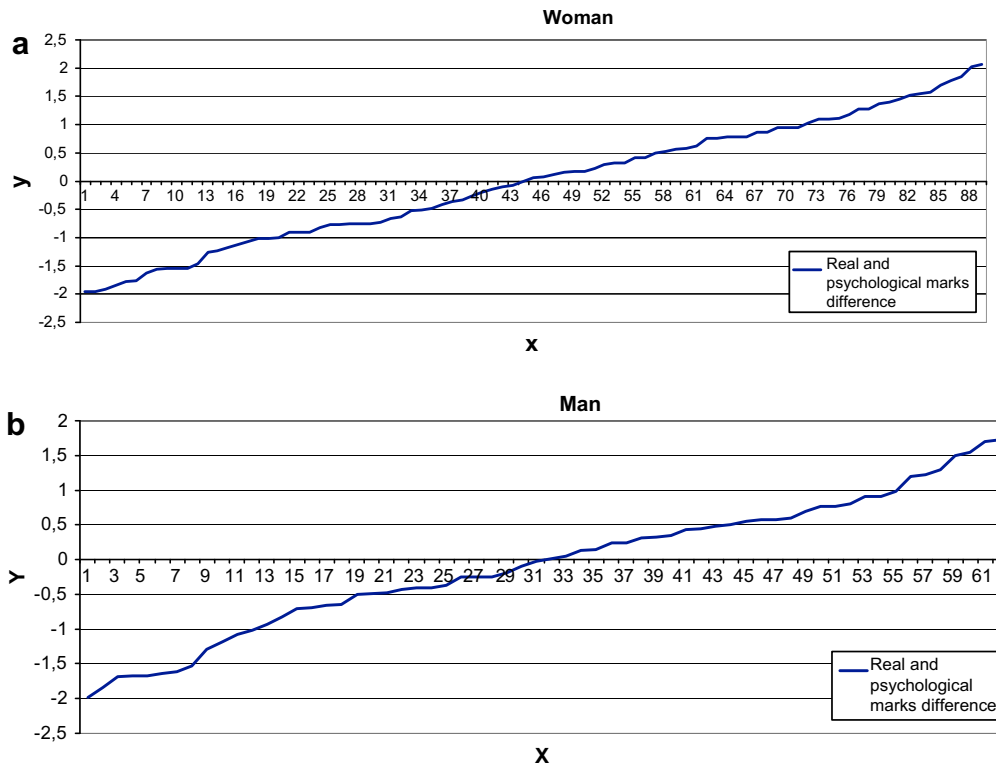


Fig. 5. Curve of the conformance of psychological test results with real test results by female students (a) and male students (b).

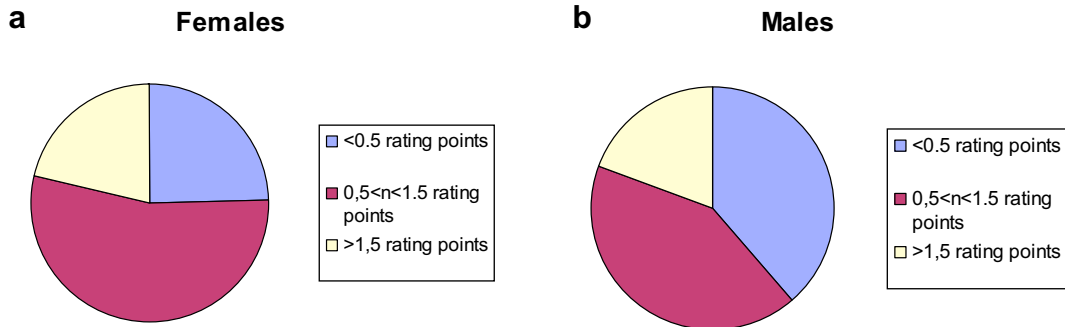


Fig. 6. Conformity of psychological test results to real test results for female students (a) and male students (b).

The farther the psychological test results distance from the *x*-axis upwards or downward, the less the results of the real test are in conformance. Nonetheless, the accuracy is sufficiently great. For example, 55 (or 61%) of the 90 female students (see Fig. 5a) had a one-point difference between the psychological and real tests. Analogically 43 (or 69%) of the 62 male students (see Fig. 5b) had a one-point difference between the psychological and real tests. It can be noticed that, in this case, the psychological testing of the male students is more effective. The average rating grade of the psychological test for male students is 8.16 and of the real test – 8.24. The average rating grade of the psychological test for female students is 8.47 and of the real test – 8.5. It can also be noticed that the difference between the psychological and real tests does not exceed 2 points for male and female students alike.

The circles in Fig. 6 illustrate rating point differences between the results of psychological and real tests of up to 0.5, from 0.5 to 1.5 and over 1.5 for female students (a) and male students (b) (Fig. 6).

A scatter plot, shown in Fig. 7, was set up for visual analysis of the interrelation between student self-evaluation and their results after testing. The *x*-axis of the diagram shows student marks calculated on the basis of their self-evaluation, and the *y*-axis shows their marks scored for the test. The obvious scatter of the observations made the visual analysis of the data noticeably more complicated; therefore, the dependencies between marks were studied using the methods of regression analysis. The analysis consisted of three phases.

The obtained data were used to set up a linear regression model. The regression equation was derived using the least squares method. The numerical expression of the regression dependence is shown in Fig. 7.

The second phase focused on the analysis of the probability distribution of the regression model residuals. It is a known fact that most regression analysis methods demand normality in the probability distribution of residuals (the assumption is that the residuals are independent random observations from a normal distribution). The modified Durbin's method (Durbin, 1961) was used to test the normality of the distribution of the residuals. The dissertation (Gribniak, 2009) describes the use of this method in detail; thus we skip the description here. The analysis of the regression model residuals revealed that their probability distribution is *normal* (with 95% probability).

The third phase included statistical evaluation of the parameters of the regression equation to test its *adequacy*. Table 1 presents the main statistics of the regression model. Part one of the table shows the significance analysis of the equation's coefficients, while part two focuses on the results obtained after the adequacy analysis of the regression model. The student's *t*-test was used to test the significance of the coefficients (i.e. to test the hypothesis that they are all equal to zero). Columns three and four of Table 1 show the calculated values of *t*-statistics and the probability for each of the coefficients to take the value of zero (the probability of obtaining a value for the test statistic that is as extreme). It should be noted that the probability for coefficients to be insignificant is negligible. The table further presents 95% confidence intervals for each coefficient. The obvious difference between the lower and upper boundaries of the obtained intervals results from the scatter of observations shown in Fig. 7. The same scatter of observations also explains the considerably low values of the coefficient

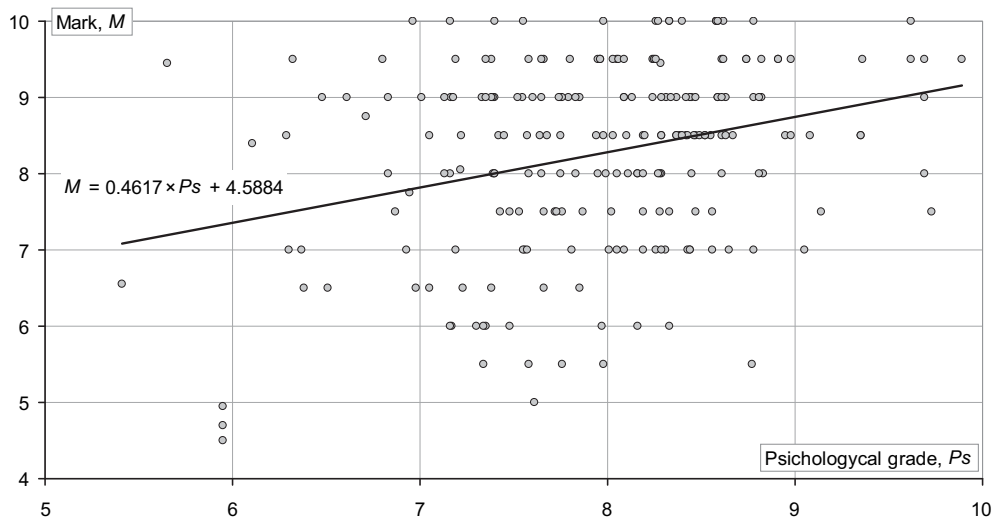


Fig. 7. The scatter plot of the student marks scored for the test and the marks based on their self-evaluation.

Table 1
The main statistical indicators of the regression model.

Coefficient	Value	t-test		95% Confidence		r	r ²	F-test	
		Value	p-value	Lower	Upper			Value	p-value
Slope	0.462	4.635	6.066 × 10 ⁻⁶	0.265	0.658	0.296	0.088	21.49	6.066 × 10 ⁻⁶
Intercept	4.588	5.768	2.661 × 10 ⁻⁸	3.021	6.156				

of regression r and the coefficient of determination r^2 , which *undeservedly* are regarded as the accuracy indicators of the regression model (Brendt, 1991: 57). The model's adequacy, on the other hand, may only be evaluated using the statistical testing of the hypothesis (null hypothesis) $H_0: r^2 = 0$, which is performed using F -test (ANOVA). The last two columns of Table 1 show the calculated values of F -statistic and the probability for the determination coefficient to take the value of zero. The results lead to a conclusion that the significance of the regression coefficients and the adequacy of the equation are unquestionable.

The analysis, therefore, reveals that the dependence between the marks scored by students for the test and the marks based on their self-evaluation is substantially justified. However, an additional investigation on the possibilities to use the suggested methodology in practical evaluation of learners on the basis of their psychological test is required.

7. Conclusions

Various methodologies (Big Five Factors and Five Factor Model, intelligence quotient tests, self-assessment) and strategies for web-based formative assessment (repeat the test, correct answers are not given, ask questions, monitor answering history, query scores, all pass and then reward strategies) are used in the effort to predict the academic motivation, achievement and academic outcomes of students. However, in the opinion of academics and practitioners alike, more than merely a student's personal attributes affects that student's advancement. Rutter and Maughan (2002) hold the opinion that the quality of a school makes a significant difference on a student's progress. The biometric system and method for psychologically assessing student progress were developed on the basis of the aforementioned methodologies, the thirteen years of experience with distance learning by the authors of this article and the technology for biometric voice analysis. The System developed by these authors are superior to the systems due to the use of biometric voice analysis and special algorithm which permits a more detailed analysis of the knowledge attained by a student.

Acknowledgement

The authors thank the referees for their valuable comments and suggestions have helped to improve this paper considerably and for the help of Ms. Vijole Arbas, who edited the text of this article.

References

- AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System*, 32(3), 407–425.
- Austin, Z., Gregory, P. A. M., & Galli, M. (2008). "I just don't know what I'm supposed to know": evaluating self-assessment skills of international pharmacy graduates in Canada. *Research in Social and Administrative Pharmacy*, 4(2), 115–124.
- Ballantine, J. A., McCourt Larres, P., & Oyelere, P. (2007). Computer usage and the validity of self-assessed computer competence among first-year business students. *Computers & Education*, 49(4), 976–990.
- Bell, A. D., Jr., Ford, W. H., & McQuiston, C. R. (December 10, 1976). *Physiological response analysis method and apparatus*. U.S. Patent. 3971034.
- Borgen, L. A., & Goodman, L. I. (1973). Voice stress analysis of anxiolytic drug effects. *Clinical Pharmacology & Therapeutics*, 19(1), 104.
- Braak, J. P. V. (2004). Domains and determinants of university students' self-perceived computer competence. *Computers & Education*, 43(2004), 299–312.
- Brendt, E. R. (1991). *The practice of econometrics: Classic and contemporary*. Massachusetts: Addison-Wesley Publishing Company.
- Brenner, M. (April, 1974). *Stage fright and Steven's law*. Paper presented at the Eastern Psychological Association Convention, New York, N.Y.
- Brockway, B. F. (1979). Situational stress and temporal changes in self-report and vocal measurements. *Nursing Research*, 28(1), 20–24.
- Cestaro, V. L. (1995) (Report No. DoDPI95-R-0002) *A comparison between decision accuracy rates obtained using the polygraph instrument and the computer voice analyzer (CVSA) in the absence of jeopardy*. Fort McClellan, AL: Department of Defense Polygraph Institute.
- Chapman, J. (1989). *VSA as a tool for eliciting confessions*. NY: Department of Criminal Justice, State University of NY at Corning.
- Clarke, E. (2005). *Psychology and law*. Juta and Co Ltd.
- Durbin, J. (1961). Some methods of constructing exact tests. *Biometrika*, 48(1–2), 41–55.
- Fitzgerald, J. T., White, C. B., & Gruppen, L. D. (2003). A longitudinal study of self-assessment accuracy. *Medical Education*, 37, 645–649.
- Formative assessment in computer supported learning. (2009, December 10). Available from: http://en.wikipedia.org/wiki/Formative_assessment.
- Gribniak, V. (2009). Shrinkage influence on tension-stiffening of concrete structures. PhD thesis, Vilnius Gediminas Technical University, Vilnius.
- Haddad, D., Walter, Sh., Ratley, R., & Smith, M. (2002). *Investigation and evaluation of voice stress analysis technology*. Rome, N.Y.
- Heisse, J. (1974). Is the microtremor usable? – the micro-muscle tremor in the voice. U.S. Congress, House Subcommittee of the Committee on Government Operations *The use of polygraphs and similar devices by federal agencies*. Hearing, 93rd Cong., 2nd Sess., June 4 and 5, 1974 (pp. 339–343). Washington: Government Printing Office.
- Hopkins, C. S., Ratley, R. J., Benincasa, D. S., & Grieco, D. S. (2005). Evaluation of voice stress analysis technology. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05)*, 1–10.
- Inbar, G. F., & Eden, G. (1976). Psychological stress evaluators: EMG correlation with voice tremor. *Biological Cybernetics*, 24(3), 165–167.
- Intelligence. (December 10, 2009). Available from: <http://www.learningandteaching.info/learning/intelligence.htm>.
- Intelligence Quotient. (2009, December 10). Intelligence Quotient – gender differences in math. Available from: <http://psychology.jrank.org/pages/341/Intelligence-Quotient.html>.
- Kaklauskas, A., Ditkevicius, R., & Gargasaitė, L. (2006). Intelligent tutoring system for real estate management. *International Journal of Strategic Property Management*, 10(2), 113–130.
- Kaklauskas, A., Kaklauskas, G., Krutinis, M., & Ditkevicius, R. (2005). Intelligence computer learning systems in VGTU. In 6th International Scientific conference, *Public relation, quality, benefits and risks* (pp. 99–104). Riga.
- Kaklauskas, A., Zavadskas, E., & Ditkevicius, R. (2006). An intelligent tutoring system for construction and real estate management master degree studies. 3rd Conference on Cooperative Design, Visualization and Engineering, Septembers 17–20, 2006, Mallorca, Spain, Book Series: *Lecture Notes in Computer Science*, 4101, 174–181.
- Kaklauskas, A., Zavadskas, E. K., Babenskās, E., Seniūt, M., Vlasenko, A., & Plakys, V. (2007). Intelligent library and tutoring system for Brita in the PuBs project, 4th International Conference on Cooperative Design, Visualization, and Engineering, September 16–20, 2007, Shanghai, People's Republic of China, Book Series: *Lecture Notes in Computer Science*, 4674, 157–166.
- Komarraju, M., Karau, S. J., & Schmeck, R. R. (2009). Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences*, 19(1), 47–52.

- Lejk, M., & Wyvill, M. (2001). The effect of the inclusion of self-assessment with peer-assessment of contributions to a group project: a quantitative study of secret and agreed assessments. *Assessment & Evaluation in Higher Education*, 26, 551–561.
- Lippold, O. (1970). Oscillations in the stretch reflex arc and the origin of the rhythmical 8–12 c/s component of the physiological tremor. *The Journal of Physiology*, 206(2), 359–382.
- Lippold, O., Redfeard, J., & Vuco, R. (1957). The rhythmical activity of groups of muscle units in the voluntary contraction of muscle. *The Journal of Physiology*, 137, 473–487.
- Marsh, H. W., Overall, J. U., & Kesler, S. P. (1979). Validity of student evaluations of instructional effectiveness: A comparison of faculty self-evaluations and evaluations by their students. *Journal of Educational Psychology*, 71(2), 149–160.
- Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26(1), 75–100.
- Mynttinen, S., Sundström, A., Vissers, J., Koivukoski, M., Hakuli, K., & Keskinen, E. (2009). Self-assessed driver competence among novice drivers – a comparison of driving test candidate assessments and examiner assessments in a Dutch and Finnish sample. *Journal of Safety Research*, 40(4), 301–309.
- Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big Five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, 93(1), 116–130.
- Papinczak, T., Young, L., Groves, M., & Haynes, M. (2007). An analysis of peer, self, and tutor assessment in problem-based learning tutorials. *Medical Teacher*, 29, 122–132.
- Ruiz, R., Legros, C., & Guell, A. (1990). Voice analysis to predict the psychological or physical state of a speaker. *Aviation, Space, and Environmental Medicine*, 63(3), 266–271.
- Rutter, M., & Maughan, B. (2002). School effectiveness findings 1979–2002. *Journal of School Psychology*, 40(6), 451–475.
- Self-assessment. (December 10, 2009). Available from: <http://en.wikipedia.org/wiki/Self-assessment>.
- Selye, H. (1951). The general adaptation syndrome. *Annual Review of Medicine*, 2, 327–342.
- Smith, G. A. (1977). Voice analysis for measurement of anxiety. *British Journal of Medical Psychology*, 50(4), 367–373.
- Sundström, A. (2008). Self-assessment of driving skill – A review from a measurement perspective. *Transportation Research Part F: Traffic Psychology and Behavior*, 11(1), 1–9.
- Sung, Y., Chang, K., Chang, T., & Yu, W. (2009). How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments. *Journal of Adolescence*, 33(1), 135–145.
- Sung, Y. T., Chang, K. E., Chiou, S. K., & Hou, H. T. (2005). The design and application of a web-based self- and peer-assessment system. *Computers & Education*, 45(2), 187–202.
- Tippett, R. (1994). *Comparative analysis study of the computer voice stress analyzer and polygraph*. University of Missouri.
- Trandis, H. (1989). The self and social behavior in differing cultural contexts. *Psychological Review*, 96(3), 506–520.
- Vainiunas, P., & Kaklauskas, A. (2005). Intelligent tutoring systems as created by VGTU. *Proceedings of 5th AECEF Symposium on Civil Engineering in the Next Decade (ASCEND), Strategies for education, research, innovation and practice* (pp. 143–148), Espoo, Finland.
- Wiggins, S. L., McCranie, M. L., & Bailey, P. (1975). Assessment of voice stress in children. *Journal of Nervous Mental Disorders*, 160(6), 402–408.
- Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education*, 11(3–4), 186–193.
- Zavadskas, E., & Kaklauskas, A. (2000). Efficiency increase in research and studies while applying up-to-date information technologies. *Journal of Civil Engineering and Management*, 6(6), 397–414.
- Zavadskas, E., Kaklauskas, A., Vainiunas, P., & Gikys, M. (2000). Property management in postgraduate Internet studies in Vilnius Gediminas Technical University. 4th Baltic Region Seminar on Engineering Education, SEP 01–03, 2000, Copenhagen, Denmark. Book Series: *Monash Engineering Education Series*, 83–86.
- Zavadskas, E. K., Kaklauskas, A., & Vlasenko, A. (2008). Web-based voice stress analyzer decision support system for e-examination. 8th International Conference on Fuzzy Logic and Intelligent Technologies in Nuclear Science (FLINS), September 21–24, 2008 Madrid, Spain, Books Series: *World Scientific Proceedings Series on Computer Engineering and Information Science*, 1, 1153–1158.
- Zimmerman, B., & Kitsantas, A. (2007). Reliability and validity of Self-Efficacy for Learning Form (SELF) scores of college students. *Zeitschrift für Psychologie*, 215(3), 157–163.