

Theses of the Doctoral (PhD) dissertation

THE MODERN METHODS OF TALENT IDENTIFICATION

dr. Balázs Klein

Supervisor:

Dr. habil. Krisztina Dajnoki

associate professor



UNIVERSITY OF DEBRECEN

Károly Ihrig Doctoral School of Management and Business

Debrecen, 2018

1. THE BACKGROUND AND THE AIMS OF THE RESEARCH AND THE INTRODUCTION OF THE RESEARCH HYPOTHESES

Workforce is an essential component of the efficiency of organizations so companies have a vital interest in identifying talented workers. My work aims to help organizations to become more effective through the improvement of enterprise practices and tools for talent identification. Within this topic I will present

- an economic overview of the concept of talent and talent identification,
- the exploration and development of new means of talent identification from an economic point of view;
- the examination and development of certain talent-identification practices of organizations.

The actuality of the topic comes from two sources: the growing importance of the human factor in the economy and the development of the methodology of talent identification.

In the course of my work I have set the goal of creating methods that enables companies to more accurately identify talent and predict performance than before and made studies with these methods to explore the nature of general intellectual ability and competences and their utilization. I am convinced that this knowledge can help business leaders develop better assessment practices and thus increase corporate efficiency.

My dissertation is based on the experiences and research of the last 25 years and through the improvement of assessment aims to help organizations to function more effectively. In order to achieve this objective I have developed and examined concrete assessment tools for examining general intelligence and competencies. In my thesis I present a matrix test suitable for measuring educative ability, a vocabulary test measuring reproductive ability and a new type of competency questionnaire showing the areas in which the efficiency of assessments and the future performance of selected candidates can be increased.

1.1. The main objectives of the thesis

The growth of the significance of the human factor and the development of talent identification methodology encourage a wide range of organizations from the business

sector to the public sector and educational organizations to introduce new talent identification methods. The ultimate goal of my work is to make better use of talent at organizational level. Although appropriate talent identification has individual benefits as well but in my present work I mainly deal with the goals organizations aim to achieve through talent identification.

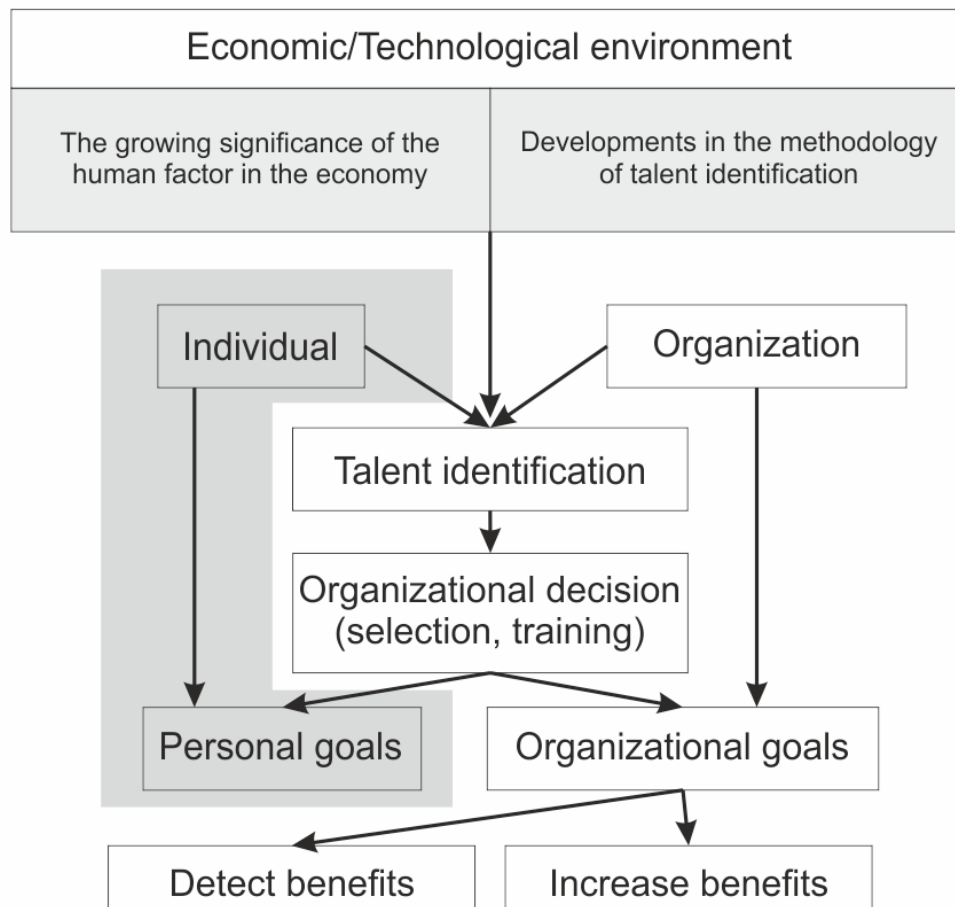


Figure 1. Personal and organizational goals related to talent identification

Source: Own work (2018)

The direct goal of my work is to detect and increase the economic benefits of using proper talent identification methods. My research model is shown in Figure 1. Talent identification is carried out by organizations so the actors of the model are the organization and the individual. The outcome of talent identification is the organizational decision. In a corporate assessment situation this decision is typically a selection decision while in an educational environment it is usually the customization of the training. The organization's decisions must always serve the organizational goals. Naturally both players have their own goals related to talent identification but in my present work I only deal with the organizational goals and the organizational benefits. These processes take place in an economic and

technological environment in which the importance of talent identification is increasing and its methodology is undergoing significant changes.

In my research I also aimed at reviewing the literature of talent identification with objective tools.

My hypotheses are related to this dual purpose. In my research I have set the goal to

- *detect* the economic benefits of talent identification
 - through an enterprise case study (H1) and
- *increase* the economic benefits of talent identification by
 - using better test tools (H2) and
 - introducing more effective organizational practice (H3, H4, H5, H6).

1.2. The primary hypothesis of the thesis

The primary hypothesis of my thesis is that improved talent identification improves organizational effectiveness.

1.3. The hypothesis of the thesis

H1. Using the new assessment tests can yield concrete monetary benefits.

H2. Modern online ability testing can be an effective alternative to conventional paper-pencil tests.

H3. Company-specific benchmarks are needed for the correct interpretation of IRT-based ability test results.

H4. Ability tests can be used as early filters in the corporate assessment protocols.

H5. Self-descriptive questionnaires can be intentionally manipulated but manipulation can be made identifiable to corporate HR.

H6. The integrated online feedback of the results is useful for the company, the test developer and the test taker alike.

2. MATERIALS AND METHODS

Over the past two decades I have completed more than 100,000 ability and competency assessments of children and adults in corporate and educational environment. My thesis gives account of some of the results of these primer data collections.

2.1. Methods of data collection and assessment

To collect data I used multiple choice competency assessment questionnaires, maximum performance tests and free text online interviewing tools.

The purpose of the test takers were varied. For example:

Selection

Candidates sought employment in large and small businesses in the private sector as well as and in institutions and organizations in the public sector. Test takers completed their tests as part of the assessment procedure required for entry.

Talent identification

Participants in talent identification programs were typically primary school children whose purpose was to get into a variety of development programs under a talent program.

Self-knowledge

Test takers with this purpose were typically university, high school or primary school students who have completed the tests and questionnaires for self-knowledge in an organized or open-ended project supervised or unsupervised.

I have developed new ability tests for my investigations in the area of eductive ability (matrix test), reproductive ability (vocabulary test). These instruments are IRT-based, responsive, online, adaptive ability tests with large item-bank. They embody significant methodological innovations such as:

- the format of the eductive and reproductive tests (binona format),
- the type of the eductive test task,
- the integrated online feedback built into the competence questionnaires and the reflective validity indicator measuring it.

2.1.1. The format of the eductive test

The Raven Progressive Matrices have been a great instrument of eductive ability - and thus of general intelligence - for nearly a hundred years. Its application proved to be extremely useful both in scientific and practical life. The long period of time elapsed did not pass over it without a trace however. Effective testing often means online testing today. For this reason a large number of online tests similar to the Raven test have been published on the Internet. These tests aimed only to renew the Raven test only in its delivery method not in its concept. The tool presented here also includes conceptual development. This development is particularly important for example in the elimination of the ceiling effect in the Advanced Raven tests. In the traditional task-format of the Raven test it was difficult to find tasks that had a good discriminatory power in the very high ability range. This is however a key issue in identifying talent. Therefore we had to find a task format that retains the core positive qualities and content focus of the Raven test but improves upon it. To this end I assumed that the basic task and the ability it aims to measure is the recognition of regularity in chaos. So I was looking for a task-format in which the original uncertainty (chaos) compared to the Raven test is increased but the regularities that need to be identified remain consistent.

In the format that I have developed the test tasks consist of 3 x 3 tiles (similarly to the Raven tests) but the task of the test taker is to select the two tiles that need to be switched so that all tiles would be logically arranged both vertically and horizontally. This format was later found to be useful in other tests as well and will be referred to as the BiNona format from merging the Latin "bin" and "nona" prefixes suggesting that we need to select two out of nine options.

After selecting the two tiles the computer automatically switches them so the test taker can see the result of the exchange. After the switch the user can "Cancel" it to restore the original status or click "Next" to accept the change and go to the next task. Typically 25 tasks are given which is typically completed within 25 minutes. There are 90 seconds to solve each task. If the time expires the system automatically moves to the next task (and submits the response). The test taker can only click on the "Next" button once the replacement has been completed. We encourage the test taker to give the answer she thinks is best even if she was unsure.

Figure 2 shows the screen after (an incorrect) swap. On the top left corner the number of the task is displayed on the top right corner the remaining time can be seen and on the bottom right corner the “Cancel” and the “Next” buttons are available.

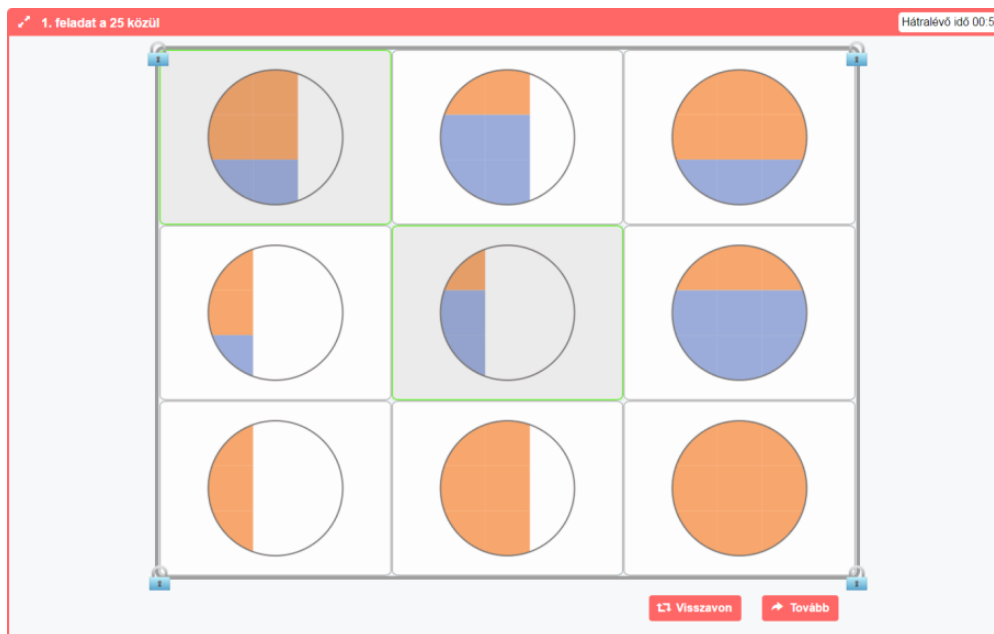


Figure 2. Personal and organizational goals related to talent identification

Source: Own work (2018)

The background of the selected tiles are grey and their contours are marked by green. As the swap has already taken place no further changes can be made in the task so the computer closed the matrix as is shown by the locks in the four corners.

This layout significantly alters the task compared to the original Raven test. Here the test taker encounters incorrect information during the rule making process. This way the process of rule making process becomes iterative: as a first step we explore and review the information that is made available to us and then derive possible rules. We use these hypothetical rules to make assumptions. These assumptions derived from the rules will then guide our attention. In this step we can either approve or reject our hypotheses based on reviewed information and create a new rule.

This new task format significantly alters the well-functioning task type. While earlier for example the difficulty of a task was often determined by the number of rules operating on a given task now the increase of the number of rules does not automatically increase the

difficulty of the task. In fact if using only one rule it becomes clear which of the two tiles need to be replaced the other rules no longer need to be recognized or taken into account.

It is my great pleasure that John Raven wrote about this test in personal communication: "This test is an exciting new development that takes forward the idea of measuring the ability to make sense out of confusion in situations where information is inherently misleading or controversial. Science – and life in general – is full of such situations."

2.1.2. The format of the vocabulary scale

The online vocabulary scale shows the words in groups of 9. The task of the test taker is to select the two words with the most similar meaning to each other. Figure 3 shows the BiNona format of the test on a standard monitor. On the top left corner the number of the task is displayed, on the top right corner the remaining time can be seen. The text of the instruction in the middle translates as “Choose the two words with the most similar meaning to each-other!”. The nine options are different Hungarian words with the chosen synonyms selected.

The screenshot shows a web interface for a vocabulary task. At the top, a red header bar contains the text "1. gyakorló feladat" on the left and "Hátralévő idő: 00 : 54" on the right. Below the header, the instruction "Válaszd ki azt a két szót, aminek egymáshoz leghasonlóbb a jelentése!" is centered. The main area displays a 3x3 grid of word options, each in a light gray box with a small icon (checkmark or minus) to its left. The words are: radiátor, reaktor, pecsétviasz, harcos, közvetítő, fűtőtest, bementő, krokodil, and méhész. The boxes for "radiátor" and "fűtőtest" are highlighted in pink, indicating they are the selected words.

<input checked="" type="checkbox"/> radiátor	<input type="checkbox"/> reaktor	<input type="checkbox"/> pecsétviasz
<input type="checkbox"/> harcos	<input type="checkbox"/> közvetítő	<input checked="" type="checkbox"/> fűtőtest
<input type="checkbox"/> bementő	<input type="checkbox"/> krokodil	<input type="checkbox"/> méhész

Figure 3. The BiNona format of the online vocabulary scale

Source: Own work (2018)

2.1.3. Integrated online feedback built into self-descriptive competency questionnaires

When reviewing and evaluating the report of the questionnaires test takers can rate and comment each sentence of the report online. This is done by the test-taker clicking on any sentence of the report displayed as a continuous text (Figure 4). The image shows a computer generated narrative description of the test taker based on his answers on a competency questionnaire.



Figure 4. The selection of the report sentences online

Source: Own work (2018)

After having clicked a sentence of the narrative description a screen is displayed which allows the selected sentence to be evaluated (Figure 5). This evaluation is done both quantitatively (using a Likert scale) and qualitatively (using short free text). On the upper part of the screen the previously selected sentence is displayed. Below that a four point scale is visible with the options “Completely disagree” (dark red), “Disagree” (light red), “Agree” (light green) and “Completely agree” (dark green). Below this is a text box where the free-text evaluation can be given e.g. “This is a sentence I really agree.”. On the bottom of the screen the two buttons are “Reset response” and “Close”.

Vélemény

"Kiegyensúlyozott képet mutat az individualizmus és a csoporttal való azonosulás terén, s bizonyára egyaránt otthonosan dolgozik csapatban vagy egyedül."

☐ Egyáltalán nem értek egyet
 ☒ Egyetértek

☐ Nem értek egyet
 ☒ Teljes mértékben egyetértek

Magyarázat:

Egy mondat, amivel teljesen egyetértek.

[500 / 39]

Figure 5. The online evaluation of the sentences of the descriptive report

Source: Own work (2018)

In my dissertation I relied on data I collected in different projects using these tools. These projects were as follows:

- More than 17,000 people have completed my online vocabulary test. Using these results I have examined among others
 - the development of reproductive ability by age,
 - the coexistence of different abilities in different ability ranges,
 - the relationship between supervised and unsupervised tests,
 - and various reliability measures of the vocabulary test.
- More than 12,000 people have completed the online educative test (SAM). Using these results I have examined among others
 - the high-stake results of over 700 jobseekers into different actual positions,
 - the relationship between the SAM test results and other variables such as the Raven Standard Progressive Matrices, socioeconomic status and competence questionnaire results for 216 students,

- the relationship between the SAM test results and the Raven Standard Progressive Matrices for 581 primary school students,
- the tangible economic benefit resulting of the test-use of 150 takers applying to an IT company,
- various indicators of reliability of the SAM test.
- In the area of competency measurement using questionnaires
 - I made an experiment with 137 people examining the possibilities and consequences of the deliberate distortion of self-descriptive questionnaires,
 - I had 637 people fill out a self-descriptive competency questionnaire and then through qualitative and quantitative methods systematically – to the level of the individual sentences of the questionnaire report – I analyzed the results and the feedback given to them,
 - I used the data given by 208 test takers to investigate to what extent is the validity of the results on a self-descriptive competency questionnaire recognizable to the test takers themselves.

2.2. Data processing methods

In high-stake situations – e.g. in selection or assessment - it is certainly not advisable to use fixed psychometric tests today. Solutions to such tests can spread over the Internet in seconds. This is even more true for tests used in an unsupervised manner. The line of defense against the spread of responses is the randomization of questions and tasks. This however provide only very limited protection - the tasks themselves remain recognizable. Test makers are therefore trying to create a unique set of tasks for everyone. To do this the test must select the current tasks from a large item bank in the background. At this point however the following problem arises immediately - if each one fills a different test how will their results remain comparable? The answer of modern psychometry to this question is item-response theory (IRT). The ability tests I present here use this measurement paradigm.

IRT is a modern paradigm of psychometry that determines how we think about making, delivering or evaluating tests. The fundamental difference between classical test theory (CTT) and IRT is that while CTT considers the whole test IRT takes into account the

different properties of each task - e.g. their difficulties. Hence the name comes from - the focus of the theory is the individual items (as opposed to the classical test theory that focuses on the whole test). The theory of IRT is based on the fact that the probability of a correct answer to a task depends on the respondent's measured ability and the attributes of the task (e.g. difficulty, discrimination power). The theory was developed in the 1950s but it only became popular in the 1970s with the spread of personal computers. The item parameters calculated by IRT method not only help us to more accurately measure the ability of the test taker but also allow the scoring to produce comparable results regardless of the individual tasks received. Having ensured that our results remain comparable it is best to give each test takers tasks that give the most accurate estimate of their ability. This is accomplished by an iterative method:

1. We give an ability estimate. If we are just starting the test then this may be the average skill level of the group.
2. From the item-bank we select the task that is appropriate for the skill level of the test taker. This will be answered correctly or incorrectly.
3. Based on the response we give a new estimate of the test takers ability and the process starts from the beginning.

Since we can optimize what task we give to the test taker we can achieve a much greater measurement accuracy for the same number of tasks or a much shorter test with the same accuracy. This method is called adaptive testing (CAT). Experience has shown that with the help of CAT technology the length of the tests can typically be reduced by half.

Figure 6 shows how the estimation of the ability of hypothetical test taker changes in the course of the test and how it influences the choice of the next task.

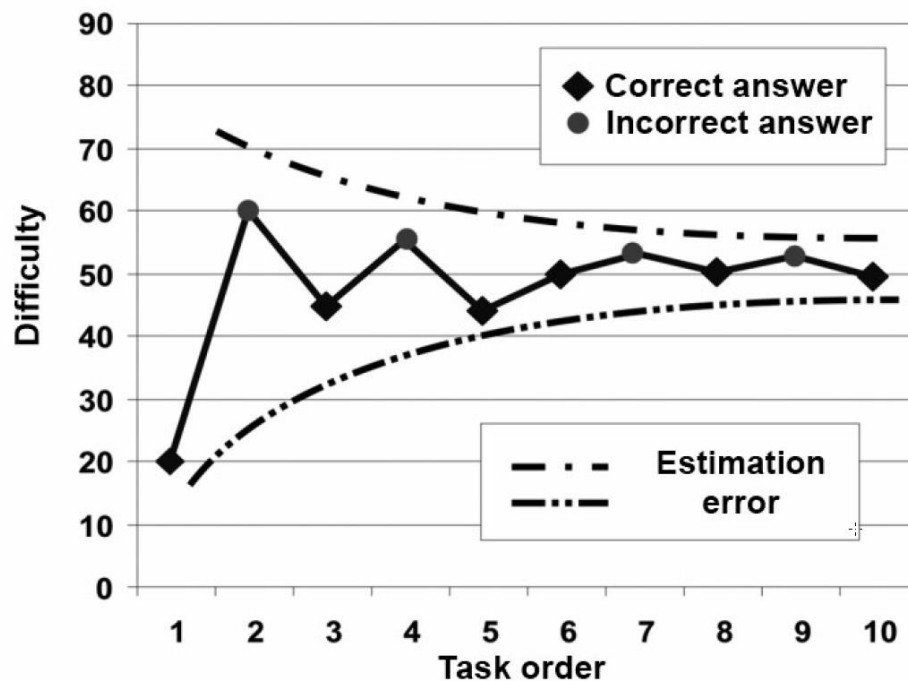


Figure 6. Adaptive testing and the precision of estimation

Source: Own work (2018)

Figure 6 shows that when encountering a correct solution the algorithm chooses a more-when encountering an incorrect solution a less difficult task. Meanwhile the estimation of the ability of the test taker becomes more and more accurate. This technology is an excellent answer to our initial problem - while it dramatically increases the efficiency of testing it also automatically ensures that individual tester takers receive different tasks.

2.3. Statistical methods

The data processing and statistical analysis of the ability and competence test results in my work was carried out using the following functions of the R statistical program:

- I used the "psych" directory of the program for descriptive statistics,
- I used Shapiro-Wilk test for normality tests,
- for T tests Student T-Tests were performed,
- In case of scatter plots I used Loeff's smoothing method,
- I created the graphs using the ggplot2 library,
- the cited correlations are Pearson product-moment correlation coefficients,
- Local gaussian correlation studies were performed using the "localgauss" library,
- bootstrap based confidence analysis was performed using the "boot" library,
- I used the "mirt" booklet for IRT models,
- linear regression calculations and their significance analysis were performed with the basic functions of R (lm, anova).

3. MAIN STATEMENTS OF THE DISSERTATION

In the following section I discuss the main hypotheses and conclusions of my work as well as its practical implications and my suggestions.

H1: *Quantifiable economic benefits may be obtained by the use of modern assessment methods.*

The cost and predictive power of the various assessment methods can be different. I have *demonstrated* that the introduction of a test method with a higher validity can have significant financial benefits at company level. My *suggestion* is that companies should review their current assessment practices and consider using more advanced methods. In order to demonstrate the economic benefits of introducing new types of assessment testing I have examined the assessment practices of two companies. I used the Bogden-Cronbach-Gleser formula to determine the actual cash benefit. Based on my results the benefits generated by the new assessment method in companies amounted to millions of forints per year in profit. My calculation significantly underestimates the actual benefits as several factors such as the risk of low performance, fluctuations and redundancies, changes in corporate strategy, the effect of aptitude test on the employer's brand, and the costs incurred by the test takers are not taken into account. Thus it can be concluded that the introduction of modern assessment tools bring significant benefits to the companies concerned. My hypothesis has therefore been *verified*.

H2: *Modern online ability tests can be an effective alternative to conventional paper-and-pencil tests.*

Assessment testing is an investment for companies. Cost plays a significant role in the return of the investment. The cost of testing without supervision at home on-line – using current market prices – is somewhat higher than the cost of paper-pencil tests. Based on my research results on-line tests provide more accurate and reliable results for job and school talent identification, the testing time can be shortened, there is no need for the test taker to travel, the online tests are able to measure the same construct as the paper pencil tests and thus their

validity is well-established while their reliability can increase over the long term (reliability of traditional paper-and-pencil tests are reduced in the long run). I demonstrated that with unattended online test use the total cost of the assessment was less than halved but most of the cost reduction occurred to the test-taker. While the use of online tests for the test takers led to radical cost reductions direct costs in the company may even become larger than those of conventional tests. Paper pencil tests are therefore cost-effective to companies as long as test takers are willing to bear their costs. Overall the company's benefits for online tests are not primarily due to the reduction in costs but to the acquisition of more talented candidates. Online tests – at current price levels - may be more cost-effective than paper-and-pencil tests if workers can choose from a number of jobs, the acquisition of talented workers is of significant importance for the company and the company's image in the labor market is important. My suggestion to organizations that use assessment testing is to consider switching from supervised paper-and-pencil ability tests to unsupervised online tests as this may have a significant benefit in many cases.

H3: Enterprise-specific benchmarks are needed to interpret even modern ability tests

The raw result of IRT-based ability tests is a Z score. This result is usually given in cumulative percentages for better understandability. According to my hypothesis in a variety of organizational environments and groups the disparity between the different groups can be so great that individual norms are needed to recognize individual differences even when using IRT based tests. In my study I compared the results of the educative test of different groups (Table 1). The test takers in the table have applied for jobs to a variety of companies and as part of the assessment procedure have completed the online SAM test.

**Table 1: The average performance of candidates
at different positions on the SAM test**

Area	Number	Avg. ability
Financial consultancy	151	1.03
IT	102	0.89
Production	24	0.69
Customer service	48	0.63
Administration	108	0.52
University	267	0.29

Source: Own work (2016), n=700

On the basis of the results we conclude that organizations with the highest average values certainly need their own norms otherwise the percentile score of the average candidate for the organization (calculated from the cumulative normal distribution function) would be 84% which would make it very difficult to identify talented candidates. For these organizations it is necessary to create their own norm which is one standard deviation above the average of the entire population.

Similar conclusion can be drawn from comparing the results of test takers of different ages on the vocabulary test. These results are shown in Figure 7. The resulting significant differences indicate that the results achieved on the ability tests – regardless of whether we used an item-response-based measurement model in our case – should be treated as raw scores and compared to well-defined norm groups.

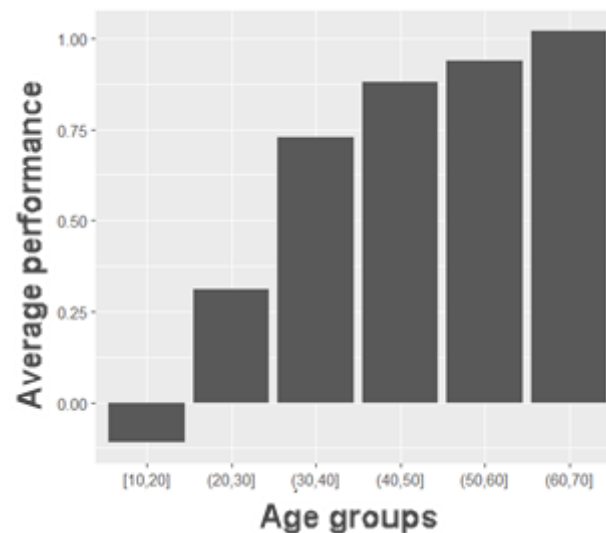


Figure 7. Adaptive testing and the precision of estimation

Source: Own work (2016, $n=13788$)

H4: Ability tests can be used as early filters in the corporate assessment protocols

Online adaptive ability tests are extremely inexpensive at the early stage of assessment and therefore – if their predictive power is adequate – can provide serious corporate benefits. According to my hypothesis low ability test results - in the jobs that require them - have a high predictive power. The differential hypothesis of intelligence states that the relationship between the different ability tests is stronger in the lower ability range. Therefore if someone performs poorly on a particular cognitive ability test he will most likely not perform well on other cognitive ability tests either. On the other hand the fact that someone performs

excellently in a particular skill area does not mean that he will achieve similarly good results in other skill areas as well. From this hypothesis it follows that rejection based on low ability test results can be a reliable form of assessment. To verify my hypothesis I had to demonstrate that the correlation between various cognitive ability tests is stronger in the lower ability ranges. Hereinafter I present the results of this research to verify this hypothesis. The results to verify the differential hypothesis of intelligence come from the Templeton project. During this project more than 17,000 young people (aged between 10 and 18) were tested. I was personally involved in the development of the test protocol in the assessment delivery and the processing of the data was done by me. As part of the Templeton project participants also filled out other tests and questionnaires besides the tests I designed: in the first round the project asked the applicants to give their age and complete 4 ability tests (not all participants completed all 4 test). To support my hypothesis I used the relationship between these ability tests.

Figure 8 shows how *local correlations* change between different ability tests in a wide range of abilities. For the first time in intelligence research I used local gaussian correlations developed originally to analyze the movement of shares and published in 2014. Based on the results the correlation is clear - with the increase in the performance level the correlations between abilities decrease.

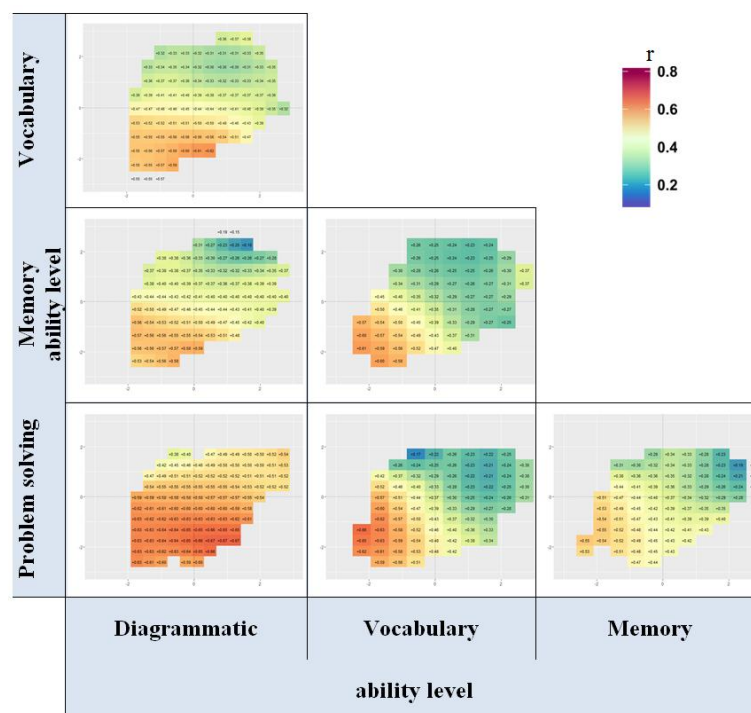


Figure 8. Local Gaussian correlations between ability tests by ability levels

Source: Own work (2016, $n=17007$)

This relationship can be further illustrated by the values obtained at the diagonal of the two-variable distributions (Figure 9).

While at low ability level ($Z = -2$) the correlation between the different ability tests was very high (0.55 - 0.65) at high ability level correlations were significantly lower (0.25 - 0.35). The difference between the local gaussian correlation values obtained in a selected low and high P1 (-1.5, -1.5) and P2 (1.5,1.5) was found to be significant in all pairwise comparisons.

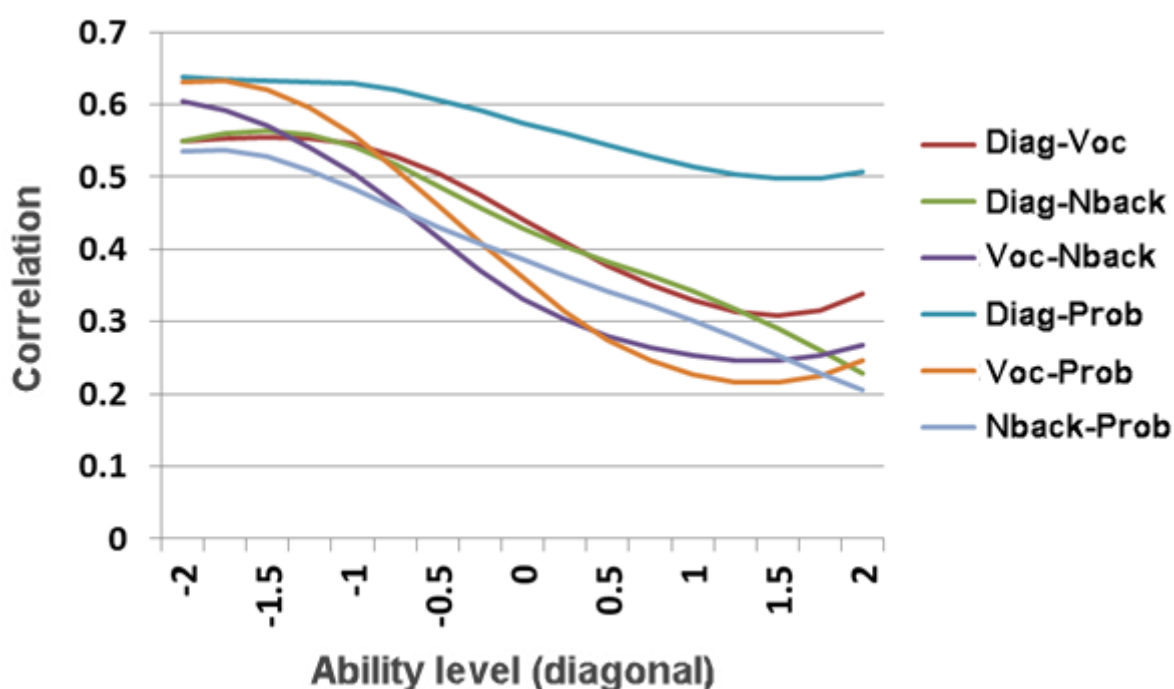


Figure 9. Local Gaussian correlations on the diagonal between ability tests

Source: Own work (2016, $n=17007$)

This significance test was performed with the confidence interval of the bootstrap method in the absence of a formal significance test method. Based on my test results when using appropriate skill tests early screening based on low ability test results can be recommended in the corporate assessment protocols.

H5: Self-descriptive questionnaires can be manipulated but deliberate manipulation can be detected

Companies often use self-descriptive questionnaires as part of their assessment procedure. To reduce the deliberate manipulation of results modern questionnaires use forced-choice form. In my study I demonstrated that test takers are able to produce a test result corresponding to a predetermined profile when filling out a questionnaire and are therefore able to deliberately manipulate the results.

Forced-choice self-descriptive questionnaires contain a built-in index to characterize the consistency of the answers. Previous studies have suggested that the consistency of answers decrease during deliberate manipulation. According to my hypothesis consistency – under the right conditions – increases during deliberate manipulation.

Based on my results I proved this hypothesis - during intentional manipulation the consistency of the answers significantly (both in statistical and in absolute terms) increased.

H6: The integrated online feedback of the results is useful for the company, the test developer and the test taker

From previous studies we know that it is possible to create personality descriptions that most people agree with. My hypothesis is that agreement with characterization - reflective validity - also depends on the validity of the characterization. I have found that agreement with the received self-descriptions monotonously decrease with the decrease of the validity of the descriptions. It follows that the test results and the reports generated are certainly and objectively linked to the reality perceived by the test takers – in other words they have reflective validity.

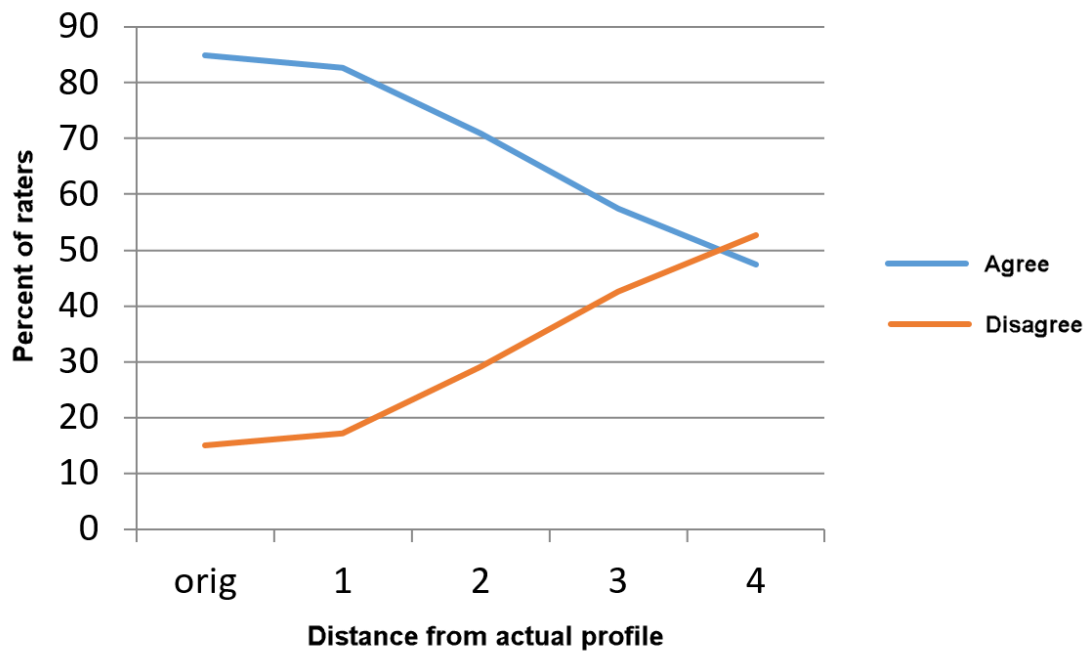


Figure 10. Agreement with self-description by reality content of description

Source: Own work (2015, $n=285$)

Part of my hypothesis is that the integrated online feedback of the results helps the development of the test tool. In my qualitative study I reviewed the feedback given to the results of a self-descriptive questionnaire and showed how the collected data helps to correct certain flaws of the instrument.

My quantitative and qualitative studies – on the whole – showed that integrated online feedback is beneficial

- to the organization since
 - it will have more and more valid information available to discuss with the candidate or to base a decision on,
 - automatically fulfilling the legal obligation of the organization saves time and money,
 - by introducing the Reflective Index I propose organizations can make a more informed choice between tools offered by providers,
- to the candidate, since
 - in some cases, it may serve as a substitute a personal feedback meeting thus saving travel time,

- automatically ensures the knowledge and the ability of authenticated amendments of the results,
- the transparency of the results ensures the quality of the instrument and the integrity of the company processes,
- the test developer since
 - it allows the quantitative identification of the weak spots of the questionnaire thus contributes to the long-term improvement of the instrument.

Using integrated online feedback therefore creates corporate value and my hypothesis has been *verified*.

4. NEW AND NOVEL RESULTS OF THE DISSERTATION

- T1:** I demonstrated that talent identification in work and education environments can be linked at device and organizational levels. The ability to measure important cognitive abilities with generic instruments that provide more accurate results than paper-and-pencil tests used to while placing persons on a single scale from elementary school age to retirement age is a major innovation that is expected to significantly change the business model of talent identification.
- T2:** I demonstrated that low scores on unsupervised online ability tests can be used by organizations as an effective pre-filter. My results clearly demonstrated that the hypothesis of declining yields is also true in cognitive performance and thus the elimination of low achievers from further assessments may be justified in many cases. For the first time in intelligence research I used local gaussian correlations originally developed to study relations in the changes of share prices.
- T3:** I have developed new methodologies and tools for general intelligence testing and have completed validating these instruments in a corporate assessment environment and demonstrated their economic utility.
- T4:** I developed and tested the methodology for the sustainable, continuous, long-term development of IRT-based ability tests. I have demonstrated that with the help of the new test methodology it is possible to create aptitude tests whose reliability increase in the long run so that both their development and their organizational application can be more advantageous than paper-pencil tests under certain labor market conditions.

- T5:** I demonstrated that the results of self-descriptive questionnaires could be deliberately manipulated under appropriate circumstances and showed how HR can detect this manipulation in assessment situations. It transpires from my study that if someone is motivated, knows the goal and the basic structure of the questionnaire and is smart enough than even the results of ipsative questionnaires can be intentionally distorted. On the other hand it has also become apparent that in the case of a high consistency result HR experts should pay close attention to checking the validity of the results during the feedback conversation.
- T6:** I have demonstrated that the use of the concept, methodology and metrics of "reflective validity" can bring significant benefits to corporate assessments. In my survey – for a specific questionnaire – I have shown that the agreement with the reports decreases monotonously with the decrease in the validity of the reports which proves that the results of the questionnaire and the reports are certainly and objectively linked to the reality perceived by the testers – in other words they have reflective validity. I have shown that measuring, using and checking this reflective validity has many advantages for all participants in the study.

5. PRACTICAL USE OF THE RESULTS

Organizations can utilize my results in the following ways:

- By showing the reliability of unsupervised online tests organizations can switch from paper-and-pencil ability tests to unsupervised online tests much greater confidence. This can have significant benefits for both the organization and the candidate in large number of cases.
- I have shown that it is possible to create a protocol that allows the continuous development of the ability tests to ensure the ever-improving quality of the test in the long run. In the future test-development companies should develop and maintain their online ability tests according to similar protocols. Consequently in the future developing ability tests will involve a larger investment and a need for continuous improvement. This can lead to the concentration of test developments over the long term and may lead to the merging of developer companies. For companies that use ability tests it is advisable to use tests that conform to this protocol ensuring that the quality of assessment remains high in the company even in the long run.
- Based on my research results it is possible to create tests that can be used to identify talents both in educational and work settings. As a result it is expedient to develop tools that work in both areas. The advantage of such tools is that in the field of education it is possible to work with extremely large numbers of test takers – which is necessary for the continuous development of the tests – while the use of tests in the world of work can have direct advantages in validating the content of educational talent projects. For organizations using tests for assessment or talent identification it seems advisable to look for tests that can demonstrate their effectiveness in both areas as this requirement automatically imposes an extremely high quality requirement on them.
- I have determined that in case of some special user groups it is essential while in other cases it is useful to use special norms even when using IRT-based tools. Test user companies should use only ability tests for which a relevant norm group has been assigned – even if the scoring process used is IRT-based.

- I have demonstrated that the correlation between the different cognitive abilities is higher at low skill levels. Consequently companies can legitimately use a recruitment procedure whereby relevant on-line skill tests are used in an unsupervised manner and allow only those who exceed a specified ability level to proceed to the further stages of the selection procedure. This protocol provides a reliable result while in most cases it provides significant savings for both the company and the candidate.
- I proved that the SAM test and Raven Progressive matrices measure similar cognitive constructs but the SAM test has many advantages over the Raven test. My research can help organizations to use the SAM test in many areas of talent testing – for which a version of the Raven tests has been used so far – and achieve significant cost savings while preserving the validity and reliability of the results.
- I have found that the use of appropriate assessment procedure yields tangible financial benefits for the companies. It is therefore advisable for companies to consider introducing them if not yet used.
- According to my findings self-descriptive competency questionnaires can be deliberately manipulated even if they are forced-choice. However this manipulation can be discovered in most cases. It is therefore advisable when making assessment decisions for corporate HR to consider the results of self-descriptive questionnaires with their inherent limitations in mind and when interpreting such results consider any extreme – low or high – value of the consistency indicator as a sign that reduces the reliability of results. In such cases it may be advisable to repeat the test or to collect information from other sources.
- I demonstrated that test takers are able to judge the validity of the results of the competency questionnaires. It is expedient to feedback and validate the results of such questionnaires with the test takers and note and consider their opinion on the outcome. It is also desirable that the authors and users of the self-descriptive questionnaires collect in a systematic way the feedback received and the conclusions drawn from the analysis of the results and use these results to improve the instruments and the decision making process.



Registry number:
Subject:

DEENK/343/2018.PL
PhD Publikációs Lista

Candidate: Balázs Klein

Neptun ID: ATP7K2

Doctoral School: Károly Ihrig Doctoral School of Management and Business

MTMT ID: 10047754

List of publications related to the dissertation

Hungarian books (3)

1. Péter-Szarka, S., Gyarmathy, É., **Klein, B.**, Kovács, K., Kövi, Z., Molnár, G., Páskuné Kiss, J., Pásztor, A.: A tehetségazonosítás folyamata, mérőeszközei és eredményei a Magyar Templeton Programban. MATEHETSZ Génius Projekt Iroda, Budapest, 68 p., 2017. (Génius Műhely ; 19.)
2. **Klein, B.**, Klein, S., Zentai, A.: Célcsoport-specifikus felmérő eszközök gyűjteménye: megváltozott munkaképességű személyek kompetenciavizsgálatára. Fogyatékos Személyek Esélyegyenlőségéért Közhasznú Nonprofit Kft., Budapest, 211 p., 2015. ISBN: 9786155610011
3. **Klein, B.**, Klein, S.: A szervezet lelke. 3. kiad., utánn., Edge 2000 Kft, Budapest, 923 p., 2012. (Segítünk, ha lehet, ISSN 1418-6586) ISBN: 9789639760226

Foreign language international book chapters (1)

4. **Klein, B.**, Klein, S., Joubert, K., Gyenis, G.: The social cage - Socio-economic status and intelligence in Hungary.
In: Uses and Abuses of Intelligence : Studies Advancing Spearman and Raven's Quest for Non-Arbitrary Metrics. Ed.: John Raven, Jean Raven, Royal Fireworks Press, New York, 568-593, 2008. ISBN: 9780898243567

Hungarian scientific articles in Hungarian journals (7)

5. **Klein, B.**: A gyerekgyár nyugdíjba megy: Az iskolarendszer elavulásáról.
Fejlesztő Pedagóg. 1 (3), 103-110, 2013. ISSN: 0866-2495.
6. **Klein, B.**: Pontosítás: A 2012-es magyar irodalom érettségi szövegértési feladat.
Új Pedagóg. Szle. 63 (1-2), 68-86, 2013. ISSN: 1215-1807.
7. **Klein, B.**, Klein, S., Joubert, K., Gyenis, G.: Intelligencia és iskolázottság Magyarországon.
Mozgó Világ. 32 (6), 51-59, 2006. ISSN: 0324-4601.
8. **Klein, B.**, Klein, S.: Alkalmasságvizsgálat.
Humánpolitikai Szle. 15 (7-8), 62-75, 2004. ISSN: 0865-7009.





9. **Klein, B.**, Klein, S.: Képességtesztek.
Humánpolitikai Szle. 15 (10), 50-63, 2004. ISSN: 0865-7009.
10. **Klein, B.**, Klein, S.: Kompetencia a munkahelyen.
Humánpolitikai Szle. 15 (5), 30-46, 2004. ISSN: 0865-7009.
11. **Klein, B.**: Tükör által homályosan.
Alk. Pszichol. 3 (1), 63-74, 2001. ISSN: 1419-872X.

Foreign language scientific articles in Hungarian journals (2)

12. **Klein, B.**, Klein, S., Nemeskéri, Z., Raven, J.: Engagement: The management of motivated employees.
Natl. Secur. Rev. 1, 158-173, 2015. ISSN: 2416-3732.
13. Gyenis, G., Joubert, K., Klein, S., **Klein, B.**: Relationship among body height, socio-economic factors and mental abilities in Hungarian conscripts.
Anthrop Közl. 45, 165-172, 2004. ISSN: 0003-5440.

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of Web of Science, Scopus and Journal Citation Report (Impact Factor) databases.

31 October, 2018

