ASSESSING SCIENTIFIC SKILLS OF APPLICANTS FOR A MASTER'S DEGREE PROGRAM

Peter Steinkellner and Karl Zehetner

FHWien University of Applied Sciences, Austria

Assessment centers have highly contributed to the quality of employee selection, training, and development in organizations. While many assessment centers are developed for business settings, they are also used as selection method in educational settings, although not as widely. Most assessment centers mainly focus on social skills. Those that simulate real problems that may be faced in the job show the highest validity. Current assessment formats for applicants of master's degree programs usually test core knowledge and social skills. However, they may underemphasize other important domains for master students, like scientific skills and the integration of scientific skills, social skills and core knowledge. We propose that for simulating real problems one may face at a master's degree program it is not sufficient to measure merely social skills. In this article we will discuss an extension of conventional assessment centers. We will show the design of an assessment center for the master's degree program 'Financial Management and Controlling' at FHWien University of Applied Sciences Vienna, which additionally measures the scientific skills of the applicants in the field of finance. We developed the following two dimensions of scientific skills: a) scientific method skills, b) argumentation skills and reflectivity. This method should help to simulate real problems one may face at a master's degree program and therefore to increase the validity of the assessment.

Keywords: Scientific skills, Assessment center, University application, Master degree.

Why Should Assessment Centers be Considered to be Transferred to Educational Settings?

In case of high demand of applicants for a university place universities are often faced with the challenge of selecting the most suitable candidates. That is not so easy, because first it has to be defined on which criteria the selection shall be based on and second, how the degree of satisfying these criteria shall be evaluated.

Assessment centers have highly contributed to the quality of employee selection, training, and development in organizations. While many assessment centers are developed for business settings, they are also used as selection method in educational settings, although not as widely. This article is dedicated to the challenges associated with the transfer of assessment centers to educational settings.

Most assessment centers mainly focus on social skills. It is crucial for validity that real problems that may be faced in the job are simulated. Assessment formats for applicants of master's degree programs usually test core knowledge and social skills. However, they may underemphasize, or even ignore, another important domain for master students: scientific skills. We propose that for simulating real

problems one may face in a master's degree program it is not sufficient to measure merely social skills, but to focus on the integration of scientific skills, social skills and core knowledge.

We will start with illustrating the problem with reference to the Austrian system of higher education, where one type of universities faces the issue of selecting applicants due to limited capacity. We have no doubt that similar conditions do exist in many other countries. We will then discuss the advantages and disadvantages of assessment centers for the selection process, with a special focus on the validity concerns that are inherent to this instrument. Then we will describe the transfer of assessment centers to educational settings at the example of a master's degree program (Financial Management and Controlling) at FHWien University of Applied Sciences Vienna. We have implemented this assessment center, which additionally to social skills also measures the scientific skills of the applicants in the field of finance, in 2014. We developed the following two dimensions of scientific skills: a) scientific method skills, and b) argumentation skills and reflectivity. This method should help to simulate real problems one may face in a master's degree program and therefore increase the validity of the assessment. Finally, we will discuss the methods for evaluating the content validity, the construct validity, and the criterion related validity of this instrument.

The Austrian System of Higher Education

In the example of Austria, three types of universities can be distinguished. The most traditional system is that of the *public universities*, with the University of Vienna celebrating its 650th anniversary in 2015. These institutions, currently 22 (Kasparovski & Wadsack-Köchl 2015, p. 17-18), are characterized by free access, there are no tuition fees at all, and, with a few exceptions, they have to accept every applicant who meets the formal requirements. This leads to a very crowded situation in many study programs at least in the first year, and to tough exams at the beginning with the intention to aggressively reduce the number of students in order to ensure acceptable student-instructor-ratios later on.

The second system is much younger; it consists of the *private universities*. The first of these now twelve institutions (Kasparovski & Wadsack-Köchl 2015, p. 36) started in 2002. Private universities do not receive any governmental funding, so tuition fees are the main means of financing in most cases, although there are exceptions where bodies like the Roman Catholic Church, the Austrian Economic Chambers, or some of the federal states of Austria provide significant funds.

The third system is that of the *universities of applied science*, which started in 1994 and consists of 21 institutions (Kasparovski & Wadsack-Köchl 2015, p. 39-40). Typically the funding is 90% by the government, and 10% by tuition fees. This system is very attractive for students, as they are instructed in small classes of maybe 40 students, and the financial burden is affordable. However, the government limits access numbers indirectly by limiting the number of subsidized places.

As an example we refer to the master's degree program 'Financial Management & Controlling', for which the described assessment center was developed. In 2014 there were 237, and in 2015 there were 241 applicants for 40 places per year. The challenge is to identify the most eligible candidates. There were many discussions on how this should be done: By a lottery? By grades acquired in the finished first cycle (the bachelor) study program? Maybe by a multiple choice test? Or with the help of a sophisticated system of assessment instruments? We decided on the latter, and so it is important to understand state-of-the-art selection methods in detail.

Employee Selection Methods in Organizations

The scientific study of employee selection is now more than a century old, as one of the first publications on employee selection was printed in 1912: Münsterberg's *Psychology and Industrial Efficiency*, which described for example the selection of streetcar motormen and telephone switchboard operators (Ryan & Ployhart 2014, p. 20.1; Münsterberg 1912; 1913). A vast number of studies, books, handbooks, and reviews on the topic have been published since then and this research has highly contributed to the quality of employee selection in organizations (Ryan & Ployhart 2014).

Many selection methods have been developed over the years. The most commonly used devices include interviews, written tests and performance-simulation tests (Robbins, DeCenzo & Coulter 2013, p. 192; Czerny & Steinkellner 2012, p. 375ff).

The *interview* is an almost universal selection method, which can be a reliable and valid selection tool, but too often it is not. The typical interview in which applicants are asked varying and often random questions in a non-standardized setting does not deliver valuable and valid information. Interviews have to be structured and standardized to be effective predictors (Robbins, DeCenzo & Coulter 2013, p. 192ff; Ryan & Ployhart 2014, p. 20.8; Czerny & Steinkellner 2012, p. 378ff).

Written tests include tests of intelligence, ability, interest, and personality, and can be moderately valid predictors for many unskilled and semiskilled operative jobs in an industrial organization. But the most important and enduring criticism of these tests is that intelligence and other tested characteristics can be somewhat removed from the actual job characteristics and from the actual performance on the job (Robbins, DeCenzo & Coulter 2013, p. 192ff; Czerny & Steinkellner 2012, p. 381ff).

This criticism has led to an increasing interest in *performance-simulation tests*, which are defined as "selection devices based on actual job behavior" (Robbins, DeCenzo & Coulter 2013, p. 193). These tests are based on job analysis data, and applicants "have to engage in specific behaviors necessary for performing the job successfully" (DeCenzo & Robbins 2010, p. 160). Therefore they meet the requirement of job-relatedness more easily than other methods (written tests, etc.) and tend to be a better predictor of performance because they evaluate the actual job behavior. The two most frequently used performance-simulation tests are work sampling tests and assessment centers. Work sampling are a miniature replica of the job and this tests are suited to persons applying for routine jobs, whereas assessment centers simulate real problems one may face on the job and are suited to managerial personnel and to applicants for highly skilled jobs (Robbins, DeCenzo & Coulter 2013, p. 193; DeCenzo & Robbins 2010, p. 160f; Robbins, Judge & Campbell 2010, p. 490; Spychalski, Quinones, Gaugler & Pohley 1997; Woodruffe 2000; Schettler 2010; Lievens 2002; Schleicher, Mayes, Day & Riggio 2002).

Assessment centers are a more elaborate set of performance simulation tests, specifically designed to evaluate a candidate's (managerial) potential. They can be defined as "a facility where performance simulation tests are administered. These include a series of exercises used for selection, development, and performance appraisals" (DeCenzo & Robbins 2010, p. 160). These exercises (leaderless group discussion, in-basket, business game, etc.) are designed to simulate the work of managers and/or experts. Assessment centers tend to be accurate predictors of later job performance (DeCenzo & Robbins 2010, p. 160f; Cascio & Aguinis 2011, p. 299ff; Ryan & Ployhart 2014, p. 20.8f; Thornton & Gibbons 2009).

One of the most popular types of an assessment center exercise is *the leaderless group discussion* (LGD). A group of participants is asked to carry on discussion about a job-related topic without a formal leader. Raters do not participate in the discussion, but observe and rate the performance of each participant. This exercise has a high face-validity when it is about a job-related topic. The reliability of the method is reasonable and is "likely to be enhanced to the extent that LGD behaviors simply are described rather than evaluated in terms of presumed underlying personality characteristics" (Cascio & Aguinis 2011, p. 300)

Overall *reliability and validity* of assessment centers are satisfactory. Interrater reliabilities vary across studies from a median of 0.60 to reliabilities of over 0.95. Candidate perceptions of assessment center exercises as highly job-related lead to a high face-validity. Reviews of the predictive validity of assessment center ratings and performance have been positive, which means that the criterion-related validity is established. The two most important points for the validity of assessment centers seem to be that assessors receive proper training and that the dimensions underlying the assessment are defined clearly (Cascio & Aguinis 2011, p. 299ff; Thornton & Gibbons 2009).

But there are also two potential problems with assessment centers regarding their validity (DeCenzo & Robbins 2010, p.160f; Cascio & Aguinis, 2011, p. 299ff; Epstein & Hundert 2002; Thornton & Gibbons 2009): The validity depends on the job relatedness of the dimensions and exercises, and assessment centers have usually low construct validity. We will discuss the low construct validity in the section "limitations and future research" and will focus here on the job-relatedness of dimensions.

Assessment centers which simulate real problems one may face at the job show the highest validity. But the main focus of assessment center exercises are often social skills where the exercises to measure them often have a low level of job-relatedness. For example the topic of leaderless group discussions in many assessment centers is often a general topic (e.g. pollution, ecological issues) and has no connection to the job the applicant applies for.

So our main conclusion is that job-relatedness and the use of performance simulation tests as assessment centers with exercises like the leaderless group discussion are the keystones of valid employee selection. Therefore selection methods should simulate real problems with job-related exercises. The challenge is: How can these findings be transferred to educational settings?

The European Definition of a Master's Degree Level

University systems have ancient traditions, and so it is not surprising that all European countries had their own systems of higher education which had been developed over centuries and therefore significantly differed from each other. In order to improve students' mobility (and consequently also professors' and instructors' mobility), these systems were aligned by defining a standard for the basic structure of study programs. European standards usually are named after the conference places where these standards have been negotiated (e.g., if non-Europeans want to visit the EU, they might need a Schengen visa, named after a small wine-making village in the Grand Duchy of Luxembourg, where the agreement on the lifting of border controls between participating European countries had been signed). So for Europeans it is natural to use the terms '*Bologna system*' when they talk about the structure of study programs, and '*Dublin descriptors*' when they discuss the competences that are to be achieved by the students in a program of a certain level.

According to the Bologna system, all countries now have a three cycle (Bachelor-Master-PhD) system, and the Dublin descriptors indicate what the acquired qualifications for each cycle should be. For a master's degree program, five competencies have been defined (Bologna Working Group on Qualifications Frameworks 2005, p. 67-68):

Qualifications that signify completion of the second cycle are awarded to students who:

1. have demonstrated knowledge and understanding that is founded upon and extends and/or enhances that typically associated with the first cycle, and that provides a basis or opportunity for originality in developing and/or applying ideas, often within a research context;
2. can apply their knowledge and understanding, and problem solving abilities in new or unfamiliar environments within broader (or multidisciplinary) contexts related to their field of study;
3. have the ability to integrate knowledge and handle complexity, and formulate judgements with incomplete or limited information, but that include reflecting on social and ethical responsibilities linked to the application of their knowledge and judgements;
4. can communicate their conclusions, and the knowledge and rationale underpinning these, to specialist and nonspecialist audiences clearly and unambiguously;
5. have the learning skills to allow them to continue to study in a manner that may be largely self-directed or autonomous.

While the entry requirement for the ability to achieve the first of these listed qualifications refers to the knowledge and understanding acquired in the first cycle (the bachelor level), the entry requirement for the other four qualifications address further competencies that may have been partly acquired in the first level, but are also influenced by other factors like professional experience, commitment, or even personality traits.

Specification of the Master's Degree Level

It is typical for study programs at universities of applied sciences that they are not oriented to disciplines, but rather to specific professional profiles. As one of the first steps in designing a new master's degree

program it is necessary to specify the general qualifications depicted in the Dublin descriptors, which leads to what we call *meta-competencies* required for a specific professional profile.

The master's degree program 'Financial Management & Controlling' at FHWien University of Applied Sciences Vienna is dedicated to the education of leading managers or executives responsible for the whole financial tasks of an organization, e.g., a CFO. Extensive research was done to identify the professional, social, communicative, intellectual and emotional competencies that are required for this position, including interviews with 124 practising CFOs in Austria (Zehetner, Pichler & Trappl 2012; Zehetner, Fahrngruber, Pichler & Trappl 2013). Finally it was possible to translate the general qualifications depicted in the Dublin descriptors into specific meta-compentencies for leading financial managers at master's degree level.

Applying the Dublin descriptors for the second cycle to the profession of a CFO leads to the following reqirements of meta-competencies (Zehetner 2013):

1. Mastery of all financial, accounting, and tax disciplines.
2. Ability to lead a team of experts and ensure fulfillment of the technical requirements, but also compliance, interdisciplinary understanding of strategy and organization, as well as fields of knowledge with management touch points such as psychology, sociology, philosophy, ethics.
3. Social competence, emotional intelligence, understanding of ethical problems, emotional stability, confidence in emotion and intuition, experience in the industry.
4. Comprehensive communication skills towards shareholders, customers, top management, media and other public.
5. Ability to independently generate new knowledge, including knowledge of scientific methods.

Definition of Meta-Labels and Deduction of Assessment Instruments

The next step was the translation of the five *meta-competencies* into *meta-labels*. We decided that the second meta-competency might be best fostered by professional experience and by the commitment of the applicants, so we determined six meta-labels to be tested:

- Knowledge and understanding of the relevant disciplines
- Professional experience
- Commitment
- Social competence
- Communicative competence
- Scientific competence

Since 2014, we start our annual selection process with a first screening in form of a computer-based multiple-choice test, where we assess the applicants' knowledge and understanding in the disciplines of finance, accounting, and tax management. This is the first funnel, where we select the 72 best candidates out of the total amount of applicants (237 in 2014 and 241 in 2015).

The next step consists of two parts, an interview and an assessment center. With the interview (15 minutes for each candidate) we assess two meta-labels, professional experience and commitment, by the applicants' answers to our questions and a third meta-label, communicative competence, by observation.

With the assessment center we assess both the applicants' social and scientific competencies. Most important in the design of the assessment situation is to ensure job relatedness. The topics on which the candidates have to work and discuss incorporate knowledge in the discipline and reflect the requirements for the profession of the CFO. The candidates are observed by assessors who have a psychology background (for the social competence) or a scientific background (for the scientific competence).

The scores achieved by the candidates in all six meta-labels (and so in all three instruments; multiple-choice test, interview, assessment center) are processed to a ranking of the candidates by using a statistical method developed by the University of Vienna's 'Department of Applied Psychology: Work, Education and Economy'. The candidates ranked 1 through 40 are accepted for the program, and a few

more are put on a waiting list in case some of the successful applicants change their mind and revoke their application after the assessment.

We want to close with a discussion of the validity of this assessment method.

Limitations and Future Research

The most important limitation of this paper and of the selection method is that at the moment we have only insufficient information about the *validity* of the procedure. As we have shown in this paper, the procedure has a high level of *face validity* because the design of the assessment procedure ensures job relatedness and the topics on which the candidates have to work and discuss incorporate knowledge in the discipline and reflect the requirements for the profession of the CFO.

There are numerous methods available for evaluating validity, Standards for Educational and Psychological Measurement (AERA, APA & NMCA 1999; Cascio & Aguinis 2011, p. 142ff) describes three principal strategies: content-related evidence, criterion-related evidence, and construct-related evidence, which match to three different specific types of validity: content validity, criterion-related validity, and construct validity. We will now discuss these different types of validity of the selection procedure.

Content validity is the degree to which the dimensions, questions, and exercises of the procedure represent the job. It is the only basic foundation for all kinds of validity. We can assume that this selection procedure has a satisfying content validity, as the dimensions are based on the Dublin descriptors and on the 2012 CFO study (124 interviews with Austrian CFOs) as we showed in this paper.

Construct validity is the degree to which a specific measured dimension (e.g., scientific competence) relates to successful performance. We will measure the construct validity using confirmatory factor analysis and other statistical methods with the data of two cohorts (applicants of 2014 and 2015) as the next step of this project in the year 2015.

Criterion-related validity is the degree to which a selection method accurately predicts the level of performance on some relevant criterion. In 2016 we will be able to measure criterion-related validity since in 2016 the first applicants who participated in this new assessment procedure will graduate. We will test whether the total score and the dimensions of the instrument accurately predict relevant criteria as academic success, marks, drop-out, and master thesis quality.

Based on these, we believe that the instrument presented here has laid a useful foundation to improve the assessment of master applicants and to increase the validity of such assessment not only for master's degree programs in finance, but in an adapted version also for master's degree programs in other fields.

References

1.   AERA, APA & NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

2.   Bologna Working Group on Qualifications Frameworks (2005). *A Framework for Qualifications of the European Higher Education Area*. Copenhagen: Ministry of Science, Technology and Innovation.

3.   Czerny, E. & Steinkellner, P. (2012). Entwicklungsorientierte Methoden des Personalmanagements. In: Heimerl & Sichler (ed.), *Strategie Organisation Personal Führung*, Wien: Facultas, 373-438.

4.   DeCenzo, D. A & Robbins, S. (2010). *Human Resource Management*. Hoboken, NJ: Wiley.

5.   Cascio, W. F & Aguinis, H. (2011). *Applied Psychology in Human Resource Management*. Upper Sadle River, NJ: Pearson.

6.   DeCenzo, D. A Cascio & Robbins, S. (2010). *Human Resource Management*. Hoboken, NJ: Wiley.

7.   Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *Jama*, 287(2), 226-235.

8.   Kasparovski, H., & Wadsack-Köchl, I. (2015). *Higher Education in Austria*. Vienna: Federal Ministry of Science, Research and Economy.

9.   Lievens, F. (2002). Trying to understand the different pieces of the construct validity puzzle of assessment centers: an examination of assessor and assessee effects. *Journal of Applied Psychology*, 87(4), 675.

10. Münsterberg H. (1912). *Psychologie und das Wirtschaftsleben. Ein Beitrag zur angewandten Experimental-Psychologie.* Leipzig: Barth.

11. Münsterberg H. (1913). *Psychology and Industrial Efficiency*. Boston, MA/New York: Houghton Mifflin.

12. Robbins, S, DeCenzo, D. A & Coulter, M. (2013). *Fundamentals of Management. Essential Concepts and Application.* Harlow: Pearson.

13. Robbins, S, Judge, T. A. & Campell, T. T. (2010). *Organizational Behavior. Essential Concepts and Applications*, Harlow: Pearson.

14. Ryan, A. M., & Ployhart, R. E. (2014). A century of selection. *Annual review of psychology*, 65, 693-717.

15. Schettler, J. (2002). Building Bench Strength. *Training*, 39(6), 54-58.

16. Schleicher, D. J., Day, D. V., Mayes, B. T., & Riggio, R. E. (2002). A new frame for frame-of-reference training: Enhancing the construct validity of assessment centers. *Journal of Applied Psychology*, 87(4), 735.

17. Spychalski, A. C., Quiñones, M. A., Gaugler, B. B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. *Personnel Psychology*, 50(1), 71-90.

18. Thornton, G. C., & Gibbons, A. M. (2009). Validity of assessment centers for personnel selection. *Human Resource Management Review*, 19(3), 169-187.

19. Woodruffe, C. (2000). *Development and assessment centres: identifying and assessing competence.* London: CIPD Publishing.

20. Zehetner, K. (2013). Was ist und wie erreicht man Financial Leadership? In: Zehetner (ed.), *Financial Leadership - Vom Experten zum CFO*, Wien: Linde, 23-32.

21. Zehetner, K., Fahrngruber, B., Pichler, R., Trappl, S. (2013). Financial Leadership - Transforming Financial Experts into CFOs. *International Journal of Business and Management Studies*, 2(2): 615-621.

22. Zehetner, K., Pichler; R., Trappl, S. (2012). Educating the Ones Who Count. *Journal of Teaching and Education*, 1(2): 379-383.