Kari Härkönen

PERSONNEL ANALYSIS AND ITS RELIABILITY

JYVÄSKYLÄN YLIOPISTO
TIETOJENKÄSITTELYTIETEIDEN LAITOS
2014

ABSTRACT

Härkönen, Kari
Personnel Analysis and its Reliability
Jyväskylä: University of Jyväskylä, 2014, 53 p.
Cognitive Science, Master's Thesis
Supervisor(s): Kujala, Tuomo

This study investigates the reliability of job candidate's personnel evaluation reports written by Recruitment and Personnel Evaluation Agency MindFit Consulting Ltd. The purpose is to clarify how similarly different readers understand the content of the reports. In the study, the questionnaire form is created to be used for scoring MindFit's evaluation reports called MindFit Analyses. The consistency of scores given by readers are then analyzed in order to get inter-rater reliability amongst readers.

Analyses investigated here are written by personnel analyst who have interviewed and tested all the job candidates. The verbal analyses are operationalized into a questionnaire for the readers to give numeric rates. The hypothesis about the different types of competences, practical and personal, is also investigated based on literature and the results of the study.

The measures using Intraclass Correlation Coefficient function in statistical program show that the raters gave significantly similar values for the most of the Criteria, especially for the practical ones, such as Technical skills and Presentation skills (consistency values 0,847 and 0,908, respectively), but also rised up the challenge of evaluating personal oriented qualities, such as Personal skills (consistency 0,695) and Orientation (consistency 0,365), which was the only quality significantly below the desired level of 0,7. The questionnaire form created in the study can be used as a base tool for the validation study.

Keywords: recruitment, personnel evaluation, competence mapping, questionnaire, reliability

TIIVISTELMÄ

Härkönen, Kari
Personnel Analysis and its Reliability
Jyväskylä: Jyväskylän yliopisto, 2014, 53 s.
Kognitiotiede, pro gradu -tutkielma
Ohjaaja(t): Kujala, Tuomo

Tässä tutkimuksessa selvitetään rekrytointi- ja henkilöarviointiyritys MindFit Consulting Oy:n työnhakijoista tekemien henkilöanalyysiraporttien, MindFitin henkilöanalyysien, luotettavuutta. Tarkoitus on tutkia, kuinka yhdenmukaisesti analyysien lukijat ymmärtävät analyysien sisällön. Tutkimuksessa luodaan kyselylomake, jonka avulla eri analyysiraportin lukijat pisteyttävät analyysien sisällön. Lukijoiden antamien pisteiden yhdenmukaisuus mitataan analyysien luotettavuuden (inter-rater reliability) selvittämiseksi.

Tutkittavat henkilöanalyysit ovat sanallisia raportteja työnhakijoista. Käytännössä analyysiteksti operationalisoidaan kriteereiksi, jotka voidaan pisteyttää numeerisesti kyselylomakkeella. Tutkimuksessa selvitetään käytännöllisten ja henkilöominaisuuksiin pohjautuvien kompetenssien arvioinnin eroon liittyvää hypoteesia saatujen tuloksien ja kirjallisuuden pohjalta.

Intraclass Correlation Coefficient -testi kertoo merkittävästä yhdenmukaisuudesta eri analyysien pisteyttäjien kesken, erityisesti käytännön kompetenssien kohdalla (esim. tekniset kyvyt ja presentaatiokyky saivat yhdenmukaisuusarvot 0,848 ja 0,908). Toisaalta ne nostavat esille henkilöominaisuuksiin liittyen kompetenssien arvioinnin vaativuuden, vastaavien arvojen ollessa 0,695 henkilöominaisuuksille ja 0,365 orientaatiolle. Tutkimuksessa luotu kyselylomake toimii tarvittaessa pohjana jatkotutkimukselle analyysien validiteetin selvittämiseksi.

Asiasanat: rekrytointi, henkilöarviointi, kompetenssi, kyselytutkimus, reliabiliteetti

FIGURES

TABLES

CONTENT

1 INTRODUCTION

The Personnel evaluation related to work performance can be thought as a matrix. At one side there is a time scale, an evaluation before the actual work is about to start and an evaluation after one already has worked some time at the job and in an organization. The evaluations before the employment are usually related to recruitment or personnel selection process. The evaluations of the current employees could be done at any time of one's career, for example in case there is a need to evaluate employee's skills related to the future career opportunities or promotions. On the other side there is an approach to evaluation itself, which can be done from the psychological perspective (for example personality) or from the domain specific knowledge point of view (competence), see figure 1.

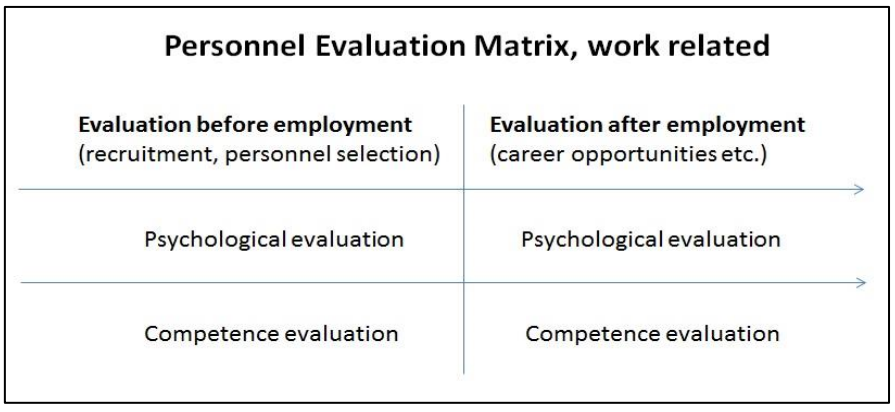

Figure 1 Personnel evaluation types as a matrix. A personnel evaluation can be executed e.g. as a part of the recruitment process or after the person has worked some time.

This study concentrates on personnel evaluation methods that MindFit Consulting Ltd (MindFit) has created for the use of recruitment and personnel selection. The approach of MindFit method is competence based including intensive tests related to technical expertise, communication, presentation and information sharing abilities. The traditional skills and competence based personnel evalua-

tion method defined by Finnish Psychology Association (FPA) is about evaluation of work performance, skills and competences using methods that evaluate the actual work behaviour. However, they don't set clear demands or regulations for the competence of the *analyst*. The analyst who uses the method and interprets the results should (according to FPA) be 'competent enough' to use such a method and evaluation process (Honkanen & Nyman, 2001.).

MindFit's evaluation covers personal interview, competence mapping, presentation test and in most of the cases also a short video presentation session. With such methods the skills and competences needed at various roles, such as development, specialist tasks and management level jobs are to be clarified. Also other personal qualities useful to be known by the employer while considering hiring someone will be clarified. The outcome of the evaluation session is a MindFit Analysis report which is the main subject of study in this thesis. MindFit methods differ from the traditional competence methods by in-depth to which the competences are to be clarified. There's also significant difference in the background of the analysts since MindFit analysts have been trained to industry field where the job candidates work.

This study works as a base for the larger study to be done in the Doctoral Thesis phase. The larger study investigates how MindFit Analyses work in practice, i.e. how well they estimate the current competence of the job candidate. To make a reliable comparison between the analysis and the employee's competence, the method for collecting such information was developed in this study.

1.1 Case: MindFit Analysis

The personnel analyses to be investigated in this study come from MindFit Consulting Ltd (MindFit). MindFit has developed its own evaluation system for analysing professional and management level people. MindFit Analysis is based on candidate interview, skills tests as well as other observations during the recruitment process. In an intensive interview and testing session a candidate's skills and competences related to a specific work field and to some specific work tasks as well as work life in general will be investigated and observed. Since the usual work field in MindFit's operations is Information Technology (IT), the specific work tasks to be evaluated are often rather practical, such as programming. After meeting the person the interviewer writes an analysis report about the candidate. The analysis is a one page verbal report describing the core competences and other skills and qualities of the candidate. On a second page it usually follows a list of competences that the candidate has filled while sending the job application to MindFit. An example of the typical MindFit Analysis can be seen in an Attachment 1. The purpose of the analysis is to give the customer (hiring company) a realistic picture of the job candidate's skills and competences, as well as one's future potential related to work career and to

some specific work tasks. The analysis about the candidate's abilities and potential as an employee will be delivered to the customer before they make a decision about the recruitment. All the analyses used in this study have been written by Kari Härkönen.

1.2 Purpose of the study

This study is about investigating the reliability of MindFit Analysis. In order to be a useful tool in recruitment it is essential that every target person reading the Analysis will understand its content in the same way. The study of consistency of the rates within the readers forms the main part of this study. The study also investigates how the different type of Criteria used for evaluating the content of Analyses by independent raters gains different results in a reliability study, and why. The main division of the rated Criteria is based on practical and personal competences.

In order to measure consistency between the rates of the readers (also called raters or evaluators), i.e. how similarly they understood the content of the verbal analysis, it was decided to operationalize the text content to a Criteria of competences that could be rated with numeric scores. The questionnaire form containing the Criteria to be scored by random evaluators was developed. The consistency between the evaluators' rates were measured by investigating Inter-rater reliability using *Intraclass Correlation Coefficient (ICC)* functionality in SPSS program (Statistical Package for the Social Sciences).

The chosen Criteria in the questionnaire form were divided into two types of competences, first five representing the practical competences (such as Management skills) and the last five personal qualities (such as Orientation). Based on this division of competence types it was expected that the practical skills are easier to be evaluated during the personnel evaluation process compared to clarifying the more psychological personal qualities. Thus the Hypothesis was set up based on an assumption that the same logic would apply while evaluators interpret an analysis by scoring such competences, because the analysis text related to practical skills was expected to be written in more detail:

Hypothesis: Practical skills can be assessed more reliably in a personnel analysis than psychological qualities.

The reliability test using ICC was executed for the data containing results from four evaluators who each scored the same twenty (20) analyses by answering to the Criteria of ten competences per analysis with the rating scale 0 to 6 (0 meaning no competence at all, 6 meaning an expert level competence).

1.3 The structure of the Thesis

This thesis presents first some literature and theories which are used e.g. as a base to create the hypothesis in chapter 2, Competence Measurement. MindFit personnel evaluation methods and practices are presented in chapter 3. The main object in the study, the MindFit Analysis, that is provided using MindFit personnel evaluation methods, is also described in this chapter. Chapter 4 presents the methods used in this reliability study, and chapter 5 its results with discussion. Finally, in the Conclusions, some thoughts and criticism are handled amongst suggestions for the future development regarding the personnel evaluation methods and practices in general.

2 Competence Measurement

It is indicated by several sources over the years that the performance between employees differs a lot (e.g. Sackman, Erikson & Grant, 1968 ; Schmidt, Gast-Rosenberg, Hunter, 1980; Motowidlo, Borman & Schmit, 1997; Schmidt & Hunter, 2004). Some people just produce more than others while doing the same work tasks. As Mark Cook says: « In an ideal world, two people doing the same job under the same conditions will produce exactly the same amount, but in the real world, some employees produce more than others. » (Cook, 2004, 1).

For example, it is investigated in an IT field that the variability between programmers' effectivity, e.g. the amount of code they make in the same period of time, can be huge. For example, already back in the 1968 a famous report about the experimental study for Advanced Research Projects Agency (ARPA) resulted that within experienced programmers (the average of 7 years of experience) the best programmer coded 5 to 28 times better than the poorest one, depending on which performance criteria was under the measure – the algebra oriented debugging produced the biggest differences in effectivity (Sackman et al., 1968). Although these results were strongly critisized later (e.g. Dickey, 1981) because of the several statistical errors such as selection problems related to programmers' experience, the outcome still pointed the best programmer as at least 5 times more effective compared to the poorest one. Later studies support the idea that computer programming is a task in which differences between programmers' productivity is substantial (Schmidt et al., 1980). Moreover, studies about expert performance suggest that the level of skills are not just related to the amount of experience but more to so called *deliberate practise* which is about how the learning plan has been designed and organized, e.g. by always trying to go beyond the personal limits (Horn & Masunaga, 2006 ; Ericsson, 2006). Whatever the real percentage of programmers' variance of ability would be within today's employees or job candidates, it should be quite obvious that the amount of experience in the certain work field or within the specific work task alone (such as programming with some specific language) does not tell enough about the real ability or skill of the person.

Thus the personnel evaluation forms an important business field in today's work life. Different evaluation methods are used e.g. for selecting people to jobs and analysing current employees in order to support their career development. The personnel selection related to recruitment is a typical situation where companies search for support from the outside, usually by cooperating with some Human Resources (HR) Agency or some organization offering psychological evaluation services. However, since the productivity differences between people are much related to the actual competences (such as some technical skills, e.g. programming), it can be asked if the personnel evaluations related to recruitment should be based more on a competence criteria than an overall level psychological tests.

2.1 Skills and competences

While interviewing and testing professionals it is essential for the evaluator to get understanding the person's suitability for the job available. There are several factors to be investigated, such as general knowledge about the industry field, skills regarding the specific job tasks and abilities affecting to person's future potential. However, it is also important to know the social and communicative skills of the person, since the knowledge of the person is usually essential for the whole organization, not just for the person itself. Moreover, if the employee's general flexibility towards the work role and its tasks does not reflect the general company policy, not even the personnel with excellent skills set will reach their full potential. That may lead to a lower performance, and at the end, lower productivity. Such an organizational 'competence' has often been referred to as contextual performance (Niitamo, 2003 ; Motowidlo et al., 1997 ; Mark Cook, 2004).

2.1.1 Practical and personal criteria

While evaluating the person at work, there could be several practical indicators related to person's work performance. For example, if the productivity is measured based on some items manufactured per day or sales per week, the differences between people could be seen rather easily. If the products to be measured are simple, i.e. it is enough to see if the worker produces 50 or 100 particles per day (e.g. assembly line in a factory) or if measurement tools in a company are precisely adjusted, i.e. the amount of code can be seen in relation to its quality, the evaluation of the work performance can be rather objective. One way to define an objective criteria is to divide it as Cook (2004) does; output/production/sales, personnel criteria, deviant behavior, training grades, work samples and walk-throughs.

The person's behavior and suitability at work could also be investigated via more personal quality perspective. There are number of things to be evaluated, that cannot be interpreted objectively. The technical results and practical numbers present only part of the data. In today's organizations the success at work is also defined by how the employee generally contributes to the organization. This is often called contextual performance, i.e. a kind of background work in order to provide good things for the organization, as Niitamo (2003) puts it. Motowidlo et al. separate the contextual performance totally from the organization's core technical processes by saying that « it does maintain the broader organizational, social, and psychological environment in which the technical core must function. » Motowidlo et al. (1997, 75). They also refer to social activities, networking (skills) and other psychological factors that affect to the environment, where the technical core is functioning. (Motowidlo et al., 1997.). Related to contextual performance, Mark Cook (2004) writes about *organizational citizenship*, which already as a term describes well about what kind of qualities are used at a work place. Such contextual performance skills are not that clearly to be evaluated, and may result a very different subjective opinions depending on who is evaluating the person. It also depends on whether the evaluator comes inside or outside the organization. Also Mussel (2012) presented *curiosity* as a personal quality that would predict many important competences useful at work. All these studies bring out the idea that personal and psychological skills are good to be known, if possible, while recruiting new employees.

There are studies suggesting that psychological or personality related qualities are to be rated less reliably than more practical oriented competences. For example, Viswesvaran, Ones & Schmidt (1996) reported in their reliability study of job performance (which was based on inter-rater reliability of supervisor reviews), that communication and interpersonal competences were rated less reliably than productivity and quality. Similarly, Wohlers & London (1989) investigated middle level managers and their co-workers who gave the rates about managers' abilities. They studied how consistently different managerial competences can be rated, and found that characteristics linked to an observable behavior (such as written skills, oral presentation and energy) got more consistent results amongst raters than got characteristics that were harder to be linked to a behavior. They also found that in general, those characteristics that were felt easier to be rated by co-workers, also provided less variability (more consistency) in their rates. However, there were some exceptions regarding which characteristics felt easy to be rated compared to consistency results. For example, leadership and decisiveness were felt easy to evaluate but gained rather big variability between raters results. This could possibly be explained partly by how raters understand the characteristic to be rated, i.e. if they have understood it similarly enough (Wohlers & London, 1989.). Moreover, according to Cook (2004), the interviewer can concentrate on assessing knowledge, skills and abilities if he or she has the exact knowledge about what the job involves. Without

such a deep knowledge only the assessment about the candidate as a person is possible, which may give poor results and allow biases to process (Cook, 2004.).

The division between practical, personal, objective and subjective evaluation is not simple. The practical skills of the person, such as presentation skills (where candidate e.g. draws a picture in a presentation test), may leave a physical result (picture) about the test subject in question (e.g. description of IT architecture). The quality of the result (picture/description of IT architecture) could then be compared rather objectively to the results provided by other candidates (e.g. drawings), but still, the interpretation has a strong subjective element since the analyst makes the final evaluation subjectively anyway.

2.1.2 The predictive value of the practical and personal competence

Probably the most important information for any employer who is considering hiring a job candidate, would be to get an estimation of how this person fits to an organization and how he or she performs at a spesific work task while the actual job starts. Thus the question is: Does the job candidate's current competence predict future behaviour of the person as an employee? The next thing to consider is: Should the current competence and future potential be measured using more practical or more personal competences?

Honkanen (2005) suggests that in the competence based personnel evaluation, only the skills at the current moment are to be evaluated. By the FPA, the long term estimations about the candidates' future performance are not recommended to do at all (Honkanen & Nyman, 2001). The evaluation results may give some preference about the future effectivity of the employee, but what it really tells is about the current situation at the day of testing it. However, in the competence based evaluation the current behavior in the interview and testing situation, as well as the available statistics related to the candidate's past work history, are usually the only factors available to form an opinion about the candidate's applicability in the future work environment. One way to get understanding of the job candidate's real knowledge and possible future behavior in the real work situation is to simulate a real-like work case. For example, Dalessio (1994) investigated newly contracted insurance agents using a video test about real-like customer interactions. Videos presented sales situations and candidates had to choose their possible action in each situation from the list of options. The test gave optimistic results regarding the potential of the agent's career in the insurance industry.

Still, it can be questioned how much weight could be put for the practical competence based estimations and for the estimations based on personal qualities about the candidate's future behavior. For example, Stanovich (2004) tells about variety of cases related to comparing clinical predictions to statistical estimations about the person's behaviour. In all such cases the actuarial prediction

that is based on statistical equations beats clinician's predictions. Dawes (2005), to whom earlier studies Stanovich often refers, seems to become even more convinced about the problems of the clinical estimations, the more time he has spent investigating the subject.

Since clinical evaluation is more related to personal quality evaluations than interpreting the practical data, it could be expected to get less reliable rating results. On the other hand, actuarial and statistic based data are practical information about the candidate's performance or past behavior (such as work history) and could thus be seen to gain more reliable rating results related to the future performance as well. In order to get useful personal quality measurements e.g. related to much appreciated personal qualities such as *contextual performance* and *curiosity*, the reliability of measurements should be guaranteed somehow.

2.1.3 Silent knowledge – meaning for a competence

In an organizational level the management of the knowledge is much about what *people* know, which *tools* are used and how company *routines* work (Cummings, 2003). The knowledge of the people is essential in a personal and in an organizational level. To have persons with the needed expertise is important for the tasks to be done, and *knowing who knows what* is very important for the organization to function effectively (e.g. in order to not waste time trying to solve something that somebody else might have solved it already). The knowledge of the person is one of the key factors regarding the specialist level job tasks. However, probably the very core thing to know about the job candidate is how this person can use his or her knowledge in action, and if this person can share the knowledge to others to learn from it. Thus the knowledge of the person should be investigated in all levels, such as how much they really know and how they can share their knowledge. It is common within senior level IT professionals that they have a lot of knowledge about the specific work tasks, as well as processes required in a professional production. At the same time, their communication skills related to sharing their own knowledge are not on the same level. Still, they might bring much additional value to other members of the organization if the others are aware of their expertise and can dig out the information out of them e.g. with the right questions and by following them closely in action.

This ability to use knowledge in action is very hard or even impossible to be find out just by interviewing person very formally, or by using some psychological questionnaire. Thus, the knowledge the person has but of which he or she does not or cannot share spontaneously, should be find out in a less formal context, such as in an equal dialogue / free discussion, or in a work-scenario situation, as well as asking the proper questions. To be clear, we call all such knowledge as tacit or silent knowledge here. Eraut (2000) emphasizes the rapid

nature of the tacit knowledge – its readiness for action when needed, and considers the ways how the knowledge could possibly be clarified by a researcher. It also seems that experts have a lot of tacit knowledge compared to persons on a novice level, and even compared to professionals that have not yet reached the so called expert status. Even very experienced professionals may need to use manuals or other references to support their work tasks while the real experts can dig the needed information out of their memory e.g. by rapidly combining their episodic experience to their semantic memory – and probably in most of the cases using this tacit information without being able to describe how they did it. Moreover, while the more theoretical oriented codified knowledge may be related to situation where it is learnt, the tacit knowledge may work more flexibly in a new environment (Eraut, 2000.). Similarly, Wagner and Sternberg (1985) propose that tacit knowledge, that is not directly taught to most of us, is related to practically oriented behavior, and thus is important for the person's career development. Khumalo (2012) investigated how middle level managers facilitated tacit knowledge transfer in an organization, and found mentoring, coaching and communication skills as very important abilities for that. From those, especially coaching and communication belong to *practical* skills that are covered in this study; Communication skills belongs directly to the Criteria to be rated in the Analysis, and coaching is partly related to Presentation skills of the Criteria (coaching skills can partly be observed in a presentation test). The tacit knowledge that would express competences in the personal Criteria (such as Orientation) can not be separated that clearly.

The simulation of the work situation could offer possibility to observe how the candidate's silent knowledge affects in action, as in Dalessio's (1994) study above. Such a real-like work simulations could be done as a part of the testing session, e.g. in a form of a presentation test (as in MindFit). Although the main idea of the presentation test is to see how the job candidate manages to share information and knowledge, it may offer an analyst possibility to separate what type of information is shared spontaneously and what type of knowledge come out only while changing from one-way presentation to dialogue between the presenter and the analyst. For example, if the test scenario simulates two people who are developing something together (e.g. discussing IT architectural solutions and possibilities), it can bring forth a lot of information about the candidate's knowledge in action. In situations where the candidate can not spontaneously share his or her knowledge, the interviewer can make prompting questions or guiding comments to help the candidate to memorize and to express the knowledge. In cognitive sciences, such an internal knowledge of the person that appears in a real life comprehension and action in different situations is called tacit knowledge [1] (see also Niitamo, 2004).

[1] There are discussions amongst researchers whether the tacit knowledge is something that is not told aloud or something that *cannot* be told or otherwise is hard to be told. Still, what matters in real work life, is the knowledge that is ready to be used in action when it is needed. Michael Eraut writes very descriptively about definitions of such a

Because the nature of the tacit knowledge (as e.g. Eraut describes it), it seems to be the essential factor the more deep expertise the employee has, but, of which can be expected to be seen in practical criteria rather than in personal one. The tacit knowledge seems to be one of the most important factors influencing the ability of an expert to use his or her competences. It is also an important form of knowledge for the career development and one of the key factors for the whole organization. Thus it is an important issue to be found out while recruiting employees.

2.1.4 The hypothesis

From the personnel evaluation or selection perspective related to recruitment, both paractical and personal criteria should be evaluated. In personnel evaluation, which is based on a personal interview, tests and inventories, and where the results are not comparable to general statistical results - which is often the case in the competence based evaluation - the both, practical and personal features are to be investigated subjectively by the analyst. However, some competences in the criteria may be seen more objectively, if those are something that can be put on paper as numbers (such as work experience in years), or to be drawn on the whiteboard when the results can be compared within candidates. Such competences can be thought as practical and thus as easier to be evaluated. If so, those can be thought to give better results in a reliability test, e.g. when measuring consistency between raters' scores. This idea, as well as literature about practical versus personal qualities (Viswesvaran et al., 1996 ; Wohler & London, 1989 ; Cook, 2004) forms the hypothesis used in this study *: Practical skills can be assessed more reliably in a personnel analysis than psychological qualities.*

2.2 Competence evaluation versus psychological evaluation

One of the main concern to decide as an employer is whether the personnel evaluation should be based more on assessing competences or psychological qualities of the job candidate. At the same time, it can be asked if it is possible to bring more value to evaluation process, if the analyst has trained or experienced in the industry field which the candidate is about to work for (i.e. the analyst is familiar with the competences). The question is relevant, since usually the analyst who evaluates the job candidate is educated to psychology and/or social sciences and has mainly gained his or her work experience from HR work having no experience nor education from the field the job candidates work for (such as IT). May, Sheng, Chitiyo, Brandt & Howe (2014) investigated internal consistency and inter-rater reliability in school settings related to functional be-

knowledge and its relation to other type of knowledge, such as codified knowledge which is more an epistemological one (Eraut, 2000.).

havior of children with disabilities. They found that paraprofessionals with more experience and knowledge about the subject (disabilities) generally agreed behavioral function to greater degrees than teachers did. It was noticed that the clinical assessment relies on raters' ability to understand environmental variables maintaining problem behavior, beside their ability to obtain data (May et al., 2014.).

The antithesis related to the practical competences and to the more psychological qualities as well as to their predictive value lead to question the whole evaluation principals. For example, based on literature that questions the relevance of estimating the future behavior (e.g. Honkanen & Nyman, 2001 ; Stanovich, 2004 ; Dawes, 2005) and studies that support the competence based evaluation of the current skills and behavior (e.g. Dalessio, 1994; May et al., 2014), the efficiency of psychological evaluation for the recruitment can be questioned. Especially, according to May et al. (2014) and Cook (2004), in cases where the analyst has substance knowledge about the industry field, the evaluation can be done in a more detail level related to the competences. Regarding the studies about the actuarial prediction versus clinical estimations about the people, Dawes (2005) has spent decades of investigating the area. He emphasizes the superiority of actuarial statistics when making predictions about the people's future behavior. The statistical data can be documented better from the practical competences and performance (e.g. related to work history), while the personal qualities require more clinical or psychological observation. Thus the reliability of practical competence can be expected to be higher.

However, the studies related to contextual performance (mentioned earlier referring to Niitamo, 2003 ; Motowidlo et al., 1997 ; Cook, 2004) bring out the importance of psychological qualities to the employees' success at work in a long term. The challenge is, how objectively that can be done. Thus, it can also be asked: What are the most important competences and qualities that could be evaluated objectively enough in order to meet the reliability requirements for the personnel evaluation and for the possible estimation about the candidate's future behavior at work?

2.2.1 The difficulty and reliability of psychological evaluation

The general outcome from the literature is that both practical competences and personal qualities are essential for the employee's future performance. Although the power of the future estimations has been questioned by the literature the studies support the idea of practical competences being stronger indicators of the performance than personal ones (Dalessio, 1994). The work life is not, however, pure skills and technical performance and thus personal qualities e.g. related to contextual performance have a significant influence to a person's success at work. However, the evaluation of such personal qualities is not that simple. A remarkable reason is that such an evaluation requires clinical as-

sessment, which is always a big risk, as Dawes (2005) notes (see above). Although the clinical assessment can be either competence based or psychological (or both), it still contains risky factors because of the clinical nature of it. Also, as Viswesvaran et al. (1996) and Wohlers & London (1989), suggest, the measurement of personal qualities cannot be done as reliably as those based on practical competences. Even if the personal qualities would have an essential affect to performance and future success, what is the benefit of evaluating those if the measures cannot be done reliably? Thus, it seems to be at least safer investment to concentrate on evaluating practical competences of the job candidates. Moreover, the practical competences seem to be easier to be evaluated using competence based measures than psychological evaluation.

2.2.2 Important competences and qualities to be evaluated

As explained above, in order to decide what the most important features to be evaluated are, one has to know how much the feature explains the performance of the candidate, and if the feature can be measured objectively and reliably. As per statistics from the studies over the years, the past behavior of the person has been pointed out as better predictor than any clinical evaluation (Stanovich, 2004 ; Dawes, 2005). The past behavior, such as work history, allows the investigation of many important practical competences, such as person's skills related to core knowledge, technical abilities, communication, presentation and management – all of which are carried out in the practical Criteria of this study. On the other hand, some personal qualities such as orientation and career expectations (about the future) may not be seen that well by investigating the past. However, some personal qualities could perhaps be seen rather well e.g. by asking a reference from the past employer of the person. Such qualities could be related to risky behavior or general applicability - also covered in the Criteria. Leong, Pearce & Huang (2013) had a questionnaire based study about how scientist and practitioner orientations affected to possible career choices of students. Although they didn't confirm the results in practice (e.g. using longitude study regarding choises people really made while they entered into a work life), the results support the idea that orientation has a significant affect to career choice.

Many studies seem to favor evaluating practical competences over the more personal oriented features. If those can be measured based on historical data, the need to use psychological tests decreases. Also, according to May et al. (2014) and Cook (2004), if clinical assessment is used for evaluating such practical competences, that should be done by the analyst who has the substance knowledge from the field – again, no need to use psychologist nor psychological tests. Then, as it has been noted that some personal qualities might be good to know by the employer, the challenge has been the reliability of measuring those. However, if the past behavioral data (e.g. work history) could bring valuable information about risky behavior and general applicability, and if orien-

tation and career plans could be reported reliably using Leong et al. (2013) method, the competence based evaluation gains even more reference over the psychological one. Thus, we could easily come to a conclusion that all the mentioned measures (practical and personal related) could be done by a professional personnel analyst with a strong substance knowledge from the industry and work tasks. The meaning of general mental ability (GMA) and personal qualities such as conscientiousness, and what competences those might bring into a criteria in personnel evaluation, should be investigated further (e.g. Schmidt & Hunter, 2004 ; Witt & Burke, 2002). If those can be proved valid qualities, and if the assessment of those can be done reliably using psychological testing, perhaps the psychological evaluation can bring some *additional* or supportive value to the competence based personnel analysis.

3 MINDFIT PERSONNEL EVALUATION

It is common within companies in the Finnish industry, that Human Resources (HR) people responsible for assigning job candidates to different departments and projects inside the company, have no specific knowledge about the industry field itself. Moreover, the lack of knowledge within HR is often especially low regarding the specific job tasks and technical criteria. Thus matching the job candidate with the right department and project which are requesting such resources is often too demanding for the HR to do with good results. The lack of knowledge within HR people has been noted during the past ten years that MindFit has worked closely with companies in the IT industry.

Personnel evaluation methods in MindFit have been built around the key level understanding of the industry standards and technologies in question, as well as well-structured, but free interview and testing sessions. The purpose of such an evaluation is to provide customers, i.e. companies who are recruiting people or subcontracting new consultants, detail information about job candidates' skills, competences and suitability regarding the customer's environment and specific work tasks. Thus, the key service is to bring more quality and cost-efficiency to their human resources process.

3.1 Recruitment process

At the beginning, close relationship with the customer's organization has to be made in order to understand their business and possible future needs regarding their work resources and people. This business and organizational understanding works as a base criteria while finding and evaluating new members to work for the customer.

3.1.1 Job Advertisement

Once the assignment to recruit a person has been arrived from the customer, the process starts with creating a suitable job advertisement if no suitable candidate exists already. In the professional IT business the differences between work task categories are enormous. For example, the competences required for creating the user interface differs a lot from the programming skills required for coding the hardware interface in a so called low level. In order to create an upper level UI in the front end, the skills set may contain graphical abilities and upper level programming languages (such as Java) while the low level developers work with the machine code or use C for programming the Kernel. And this could be the difference just from the technical perspective. Other major differences in the work profile appear e.g. regarding the communication and management skills, or the knowledge of the business field and work processes.

Thus the suitable job advertisement will not just wake the interest within people in general. It also specifies the job criteria in a way that people who work in the IT field can see if the job is suitable for them and whether it is worth using their time to create an application. There are, however, controversial results related to attraction of realistic job descriptions based on *negative* information about the work place, see Bretz, Jr., R & Judge, T. (1998). This is of course not the same as to describe the level of difficultness realistically about the work tasks. Actually, it has been noted in MindFit that the best possible job advertisement regarding the professional level jobs will weed out the candidates not suitable for the job, and wakes the interest within those who are the real target group. That cannot be emphasized enough since filtering the right candidate from all the applications is very time-consuming. This overload of job applications is actually one of the main reasons why companies use recruitment agencies – they want to reduce HR costs spent for finding the right professionals from the huge amount of applicants.

3.1.2 Headhunting and partners

Beside advertising jobs MindFit has a contact network of partners that will be contacted especially in cases if the job advertisement will not produce the desired candidate base. The other way to search suitable person is to do so called headhunts. Those can be done based on customer's hints or by using own channels.

3.1.3 Competence inventory

MindFit has developed its own web-based competence inventory to gather the relevant skills and competences from job candidates. In practice, candidates can choose from the list of competences the ones that match with their background.

The competence list will be collected using the same form that candidate leaves to MindFit while applying for a job. The listed competences have been selected based on information got from the customers as well as by following the industrial markets in order to list only the most relevant ones. It has been important to not list all the possible competences in order to keep the job candidate motivated to fill the competence form. The idea is that from the too long list of competences it could be difficult for the candidate to priorize the most relevant ones, or it could even decrease the motivation to fill the list. The information collected via competence inventory includes such data as competence (e.g. programming language or project management skill), amount of work experience per each competence and the information about when the competence was practiced last time.

3.1.4 Pre-selection of job candidates

The candidates for the rest of the recruitment process are selected from the received job applications as well as from other channels, such as those resulted from the headhunt activities or partner network. Over the years MindFit has given more weight to this early selection process starting from the realistic job advertisements to early phase selection. It has been noted that by spending more time with the candidate over the email and phone on early phase, the process to evaluate suitable candidates is more effective since only the most potential ones will be selected to personal meetings and thus to be tested in a real face-to-face situation. This leaves more time to spend with potential candidates and allows MindFit to do more other productive recruitment tasks.

3.1.5 Interviewing, testing and evaluation

After the detail review of candidates' work history (e.g. CV) and pre-collected competence data as well as pre-selection by email and phone calls the most potential ones will be invited to 1 to 1,5 hour lasting combined interview and testing sessions with MindFit's personnel analyst. This will be described more detail in the Evaluation Methods.

3.1.6 Writing an analysis

Based on observations and results from the interview and testing session the MindFit analyst will write a report which is the main object of the study in this thesis. The analysis follows always the same format covering the general skills and competences needed in the industry, as well as detailed job task specific skills. See also chapter *MindFit Analysis* and an attachment 1.

3.1.7 References

One of the most appreciated selection methods in recruitment business is to verify the candidate's past work performance from some earlier employer, supervisor and/or colleague. The reason for such priorizing the meaning of references may be due to the statistical observations that past achievements can tell more about the work performance than psychological tests. For example, well known statistical observation that Stanovich (2004) has directly taken from Dawes (1994), is that: "the best predictors of future behavior are past behavior". Although the comment was not directly related to hiring persons, but behavior in general, it is noticeable since the behavior is the key interest in work performance. In MindFit, the references belong to a process as a natural part of it. However, based on experience during the past ten years, references have not significantly contradicted with the analysis made by MindFit. Thus the weight of the reference in MindFit personnel evaluation process is quite low. This can be due to the very specific competence evaluation methods regarding both technical and communicational aspects that can be cleared out in an analysis. Thus they are in line with the past employers' observations.

3.1.8 Customer delivery and follow-up

The analysis report together with the candidate's self-made CV and competence profile are to be delivered to the customer. For the past seven years almost all the candidates have been video-interviewed as well. The video session results about a two minute long presentation video from the candidate and is also available for the customer. Later, after the candidates have worked for the employer, the follow-up is to be made in order to get information about the success of the recruitment and customer satisfaction. If the candidate works as a MindFit consultant on a customer site, regular development discussions are to be made between the candidate and the MindFit analyst in order to gather information about the candidate's own satisfaction, motivation, work performance and future career hopes.

3.2 Evaluation Methods

Evaluation methods in MindFit are based on strong pre-collected data and prephase discussions over the phone and an intensive personal meeting with a deep dialogical interview and testing session. The focus in an evaluation is competence based, concentrated on the real skills, competences and capabilities of the candidate regarding the real work performance.

3.2.1 The knowledge of the analyst

The basis for successful personnel evaluation is that the analyst has been trained to psychological principals and competence based evaluation theories. From the MindFit perspective, the other important dimension is the knowledge about the work field where the evaluated candidates come from. The traditional evaluation methods, e.g. those defined by Finnish Psychological Association and Edita, are focused on the theoretical knowledge of psychology and personnel evaluation, as well as to the analyst's work experience as a personnel evaluator (Honkanen, 2005). When evaluating people, MindFit uses the combination of theories (psychology, cognitive science and other evaluation theories) and practical experience from the work field in business to make in-depth mapping of the candidate's true skills, competences and behavioral working strategies.

3.2.2 The Interview and testing session

When the applicant has been qualified to proceed for the interview, the personal meeting lasting 1 - 1,5 hours is to be arranged with him or her. During the meeting the competence data that the candidate has provided via inventory will be clarified together with the analyst. In practice, in-depth discussions and detailed examples of where and how the candidate has used such competences are to be proceeded. The interview follows a pre-defined structured agenda, but the order of the covered subjects may vary from case to case since the purpose is to create 'free' atmosphere where the candidate can open up about his or her true skills, desires and motivations. The interview is more a dialogue between the analyst and the candidate. That is especially the case if the candidate is a very experienced professional or a manager with a strong industrial background. Same kind of dialogical discussion could be done with anyone, no matter if the candidate is experienced or not. However, candidates with senior level background are more practised discussing things instead of being questioned. With this method the most deepest and complex experiences and motivations can usually be lifted up into a discussion. Beside the interview the session also covers tests, such as forced choice and presentation test described below.

3.2.3 Silent knowledge

While interviewing a job candidate, one of the most important things is to tack out the possible silent knowledge of that person. That is especially important related to very technical level job profiles and candidates who are expected to be senior level specialists or experts, since probably majority of them know more than they can tell others (see also chapter *Silent knowledge – meaning for a competence*). In a MindFit interview session the dialogue mentioned above forms an important source of information in order to get the relevant knowledge out of the candidate. Although it might be difficult or even impossible to find the proper questions for the candidate to answer specifically, the free dialogue of-

ten produce some spontaneous details about candidate's know-how, especially when the candidate faces challenges in speaking formally about skills and background. One of the main sources for understanding the candidate's true skills and competences is the presentation test, see below. From the silent knowledge point of view, there's a tremendous effect in the way how the test situation is to be accomplished. If the candidate shows hesitation or too much tension, it is important to decrease the formal aspect of the presentation in order to form kind of a dialogue or cooperation session between the analyst and the candidate.

3.2.4 Presentation test

One purpose of the presentation test is to be a formal situation in order to evaluate the candidate's ability to present things in a context where different type of audiences may be present. One of the main aspects in the test is to provide an analyst understanding about candidate's skills to communicate and to share information within different interest groups. The ideal candidate can perform technical issues in a way that the information brings value to other technical people, but at the same time presenting the subject from the business and customer perspective. Thus, the candidate should be able to communicate on different levels, in-depth, overall, and in a way that less experienced people could follow the presentation as well. Those parts of the presentation are separated by asking or guiding the candidate to explain in a more detail or upper level matters about the subject, if he or she does not consider all the perspectives by himself or herself.

The session then moves to a phase where the analyst takes a more active role in the discussion in order to form a cooperative situation that can simulate a normal work task e.g. related to architectural IT development together with the colleague. Thus the presentation session does not just provide results about the candidate's skills regarding the presentation techniques and style, it is also a very important possibility to get understanding about the tacit knowledge of the candidate. For that matter, at some point with every candidate, the session needs to go into a less formal phase where the candidate can more openly tell about his opinions and share his or her knowledge.

3.2.5 Forced choice method

During the meeting the candidate will be tested with so called forced choice verbal test where the candidate has to choose rapidly between two roles or two competence descriptions. For example, the question could be to choose between Software Tester and Developer roles or between Project Manager or Line Manager roles. The candidate has only a few seconds to consider and answer each question. To see an example of the template for collecting data with forced choice method, as well as instructions for how to do it, see attachment 4.

3.2.6 Video

The video session is a short, around two minute long recording where the candidate has to present his or her background, core skills and future plans briefly. The analyst may ask unexpected questions or give some guidance for the candidate if he or she cannot speak spontaneously about the requested subjects. The whole session comes as an unexpected situation to the candidate without a possibility to prepare speeches. To be equal system between candidates the video is always recorded as a one time shot without a possibility to try again. The presentation language is English also for the Finnish candidates, so it also gives objective information about the candidate's language skills.

3.3 MindFit Analysis

The structure of the MindFit Analysis has originally been put together with MindFit's customers during the years 2004 and 2005 mainly. With the help of customers' feedback there have been minor adjustments and development steps over the years related to the structure. Still, the structure and the central issues handled in an analysis have mostly remained the same – the development has mainly related to logical description style and the level of in-depthness regarding the specific competence evaluated, such as how deeply the *Presentation skills* of the job candidate have been described. Even though there were several analysts in MindFit between 2007 and 2013, the structure of the analysis remained the same because all the analysts followed the common principals in a testing and interviewing session, as well as when writing the analysis. Analyses are written for the customers, who read them and use them as one major tool related to recruitment decision. Since customers have took a part of developing the structure of the analysis, it is expected that it contains relevant Criteria of competences.

3.3.1 The main sections of MindFit Analysis

The example of MindFit Analysis is presented in an attachment 1. The analysis is a one page verbal written report that has been divided to three main sections, and one supportive section on a second page:

Shortly
This section gives a short cross-section about the evaluated candidate. The purpose is to give the employer impression about the candidate's overall suitability at one sight, very quickly. It lists the core issues about education, work history, core expertise, management skills, communication style, presentation skills and orientation. For the practical reasons, the recommendation for the salary and/or hour rate, are presented briefly as well. The mentioned issues are listed in a

formal analysis template and are thus handled respectively in every analysis. There may also be a short sentence about the general *profile* of the candidate, e.g. *Programming oriented technical project manager*. If there are *special* observations, such as remarkable achievements or very important certificates, or, if there can be seen noticeable risks related to the candidate's applicability (such as commitment), those can be lifted up to this short presentation section of maximum of ten rows.

Person Description
The purpose of this section is to give a deep analysis of all the competences and other issues mentioned in *Shortly* (except for the salary/price which is not analysed here), plus give other relevant information about the candidate. Thus the candidate's whole profile with core expertise, possible weaknesses and development needs are described here. All the observations that contradict with each other, such as if the candidate seems strongly oriented to work with some field or tasks without capability to do that on a professional level, are reported here. This section may also give an impression about the candidate's future potential.

Applicability
The main message for the employer regarding the candidate's applicability and suitable work roles and tasks is presented verbally here. The roles and tasks that will not suit for the candidate or that may cause adaptation risk, are presented here as well. The section may also give suggestion for the suitable roles *in a long term*, but that claim is only *an estimation* based on observation at the moment of the testing and interviewing session. To make firm conclusions about the employee's career it will always require a follow-up of the performance and orientation of the employee.

Competence List
The competence inventory was launched in 2007 when the new Customer Management System (CMS) of MindFit was taken into use. The new system allowed job candidates to insert their competences via the same web application form that they used for applying a job. The system provides a list of competences that are to be cleared out together with the candidate and the analyst, and the verified list is presented on a second page of the MindFit Analysis. This fourth section of the analysis supports the other sections mentioned above, as it gives an easy to read list about the candidate's technical and other competences.

3.3.2 The purpose of the MindFit Analysis

The clear purpose of the analysis is to give the employer information that cannot be got from anywhere else in such a detail level. The analysis is delivered beside the CV, competence list and video presentation of the candidate, and it gives an objective description that may either support or contradict with the CV or other information stated in other material. Although MindFit verifies that the

data in CV, competence list and in other material is based on real experience and performance of the candidate, the analysis report opens up the candidate's *ability* behind the experience.

4 METHODS

In order to be valuable test or questionnaire in behavioral science, the test should meet the common reliability requirements defined in the statistical and psychometric theories. In general, the reliability examines how consistent the results of several researchers are from the same phenomenon, see e.g. Järvinen & Järvinen (2004). One base principal to make a reliable measurement is to prevent errors to occur in a first place. It is far more difficult to assess the effect after the error has occured. As Nunnally & Bernstein (1994) presents:

> One reduces measurement error by (1) writing items clearly, (2) making test instructions easily understood, (3) adhering closely to the prescribed conditions for administering an instrument, (4) making subjective scoring rules as explicit as possible, and (5) training raters to do their jobs.

According to Nunnally & Bernstein (1994) the satisfactory level of reliability in the early stages of research is e.g. 0,70.

4.1 Setting up the methods

The main challenge was to create methods for clarifying if the different, randomly chosen HR/IT personnel would understand the verbally written MindFit Analysis text similarly. Instead of comparing their verbal opinions qualitatively, it was decided to choose the most relevant Criteria from the analysis text in order to be able to score each competence of the Criteria by readers using numeric values. For that, the analysis text was divided to Criteria containing the ten most important competences that could be cleared out from the analysis format. By investigating hundreds of MindFit Analyses written between 2005 and 2013, the chosen ten competences seemed to be described detailed enough for scoring. Next, the scale for scoring competences had to be chosen. The chosen scale is a seven-step rating table from value 0 to value 6. The Criteria and the rating scale form the key part of the questionnaire, see figure 2. The results

got from the evaluators i.e. scores from randomly chosen HR personnel were compared by calculating Inter-rater reliability, which is a statistical way to compare the consistency between the evaluators' rates (Coolican, 2005).

	Table 1							
	Criteria	The level of criteria (ability, skill or trait)						
	(Ability, skill, trait, appl.)	0	1	2	3	4	5	6
1	Core skill						x	
2	Technical skills						x	
3	Communication skills					x		
4	Presentation skills				x			
5	Management skills	x						
6	Personal qualities						x	
7	Orientation						x	
8	Career expectations				x			
9	Risks (0 = high; 6 = no risk)						x	
10	Applicability					x		

Figure 2 The Criteria in Questionnaire form, 10 competences were rated in a 7-step scale.

In practise, Intraclass Correlation Coefficient function was used in SPSS program.

4.1.1 Questionnaire Form

The Criteria into a questionnaire form was collected from the practical competence factors (Core skill, Technical Skills, Communication skills, Presentation skills and Management skills) and from the more personal oriented competence factors (Personal qualities, Orientation, Career expectations, Risks and Applicability) described verbally in a MindFit analysis text. Those form the Criteria in the Questionnaire form (figure 2). The first five practical competence factors are supposed to describe a person's technical and other practical expertise crucial for the successful work performance. The meaning of the latter five applicability factors is to describe the suitability of the candidate based on his or her personality and overall suitability regarding the career in the IT field.

It is also essential for the questionnaire that the Criteria would be understood the same way by every evaluator. Thus each competence was explained in detail. The explanations were visible for the evaluators while they scored the Criteria, as in an example explanation of one of the ten competences in a Criteria, see figure 3.

Criteria Descriptions:
Core skill = The very best skill over other knowledge and competence that the candidate has. For example, for the developer it could be a specific programming language. For some person the Project Management could be the very best skill to be evaluated here.

Figure 3 Criteria Description, Core skill

4.1.2 Operationalizing the Criteria

The idea was to transfer the verbal analysis text to competences that could be rated with numbers by evaluators. The ten competence factors mentioned were operationalized under the rating scale in order to transfer the content of the analysis to numeric values. It was ended up to a seven-step scale from the value 0 (meaning no skills at all) to value 6 (expert level skills), see chapter *Choosing the Rating Scale.* For example the level of the Core skill mentioned in a MindFit analysis is to be evaluated based on the candidate's practical competence level at such a work task. On the other hand, possible risks related to recruitment of the candidate (such as candidate's motivation in a long term) are to be evaluated on the same scale so that the level 0 means high risk and the level 6 refers to the situation where the Employer (evaluator) cannot see any risks at all related to recruiting the candidate. Thus higher result (level) in every evaluated Criteria implies better competence or suitability (or less risk) regarding the work assignment in IT field. As for the competences in Criteria, in order to confirm that all the evaluators using the scale would understand the meaning of each score similarly, the scores of the scale had to be described in detail as well, see example in figure 4. The complete Criteria Descriptions and Level Description per Criteria can be seen in Attachment 2.

Level Descriptions per Criteria:
Core skill, Technical skills: 0 = No competence at all; **1** = Has weak understanding of the issue. Is not able to work independently with the task; **2** = Has basic knowledge of the issue. Can help in performing the task if getting strong guidance; **3** = Has an overall level understanding of the issue. Can perform the task rather indepently if getting support when needed; **4** = Good knowledge of the issue. Can perform the task independently and search the needed knowledge from the right sources; **5** = Excellent knowledge of the issue. Can perform the task independently and tutor others; **6** = Expert of the issue. Can improve the method itself, create course material and teach others.

Figure 4 The level descriptions for the Core skills and Technical skills.

4.1.3 Choosing the Rating Scale

To define the scale for the numeric Criteria values operationalized from the MindFit Analysis, it had to be decided what kind of a scale would be suitable. The chosen scale would have affect to statistical analyses performed for the rating results in order to define the consistency between the results of the raters. If the wrong scale was chosen all the needed statistical tests related to the study would not be possible to be executed (Nummenmaa, 2004). Thus the wrong choice might have questioned the whole reliability study of the questionnaire form.

Although the competence is a continuous variable, i.e. the level of work performance in a specific competence does not vary in steps between people, rating the level of competence had to be scored on some suitable discret scale in order to get clear numeric values from the evaluators. Continuous scale is more accurate than discrete, but very difficult to implement so that evaluators would be motivated to fill the form. However, in this case the 7-step discrete scale is more closer to continuous variable compared to some discrete scale which contains 5 or less values for scoring. Thus the behavior of the data could be thought to act as it would be continuous since such a data gave more possibilities regarding the available options in statistical tests. Competences could also be measured with Interval Scale. However, for the practical reasons the scores were limited to positive values starting from zero (0) and using less accurate Ordinal Scale, as Nummenmaa (2004) writes about the variables and scales and about the practical problems regarding the need to use less accurate scale.

In order to confirm that the evaluators would understand what each value in the scale implies, values had to be explained carefully. The explanations were rather comprehensive and therefore the scale was limited to a reasonable amount of options (7). For example, if the rating scale would contain 10 options with each option having different explanation, it would be difficult for the rater to keep the meaning of each option in mind while considering what to choose. As known from the psychology, the amount of parallel chunks that a person can keep in mind varies from 3 to 9 (Hakkarainen, Lonka & Lipponen, 1999 ; Cowan, 2000 ; Vergauwe & Cowan, 2014). Although while repeting actions (e.g. filling a form) the size of one chunk in a short-term memory grows making it possible to handle more actions and to keep more options in mind. Still, the questionnaire form is to be filled once and thus such a learning effect related to enlargening units is rather limited.

4.2 Pre-Reliability Test

Before the actual reliability test was chosen the first version of the Questionnaire form was investigated by Kari Härkönen and Tuomo Kujala, who worked as evaluators in this Pre-Reliability phase. The idea was to clarify, if the two different evaluators who read the same analysis will understand the content similarly. The purpose was also to clarify whether the questionnaire is presented in a form that the filling instructions could be understood the same way. If the results for both of the questions above were positive with the acceptable error marginal (good reliability), the form could be tested with the more time consuming and official reliability test, which in fact forms the main subject for this study – i.e. this investigation worked as a pre-reliability test.

4.2.1 Pre-Reliability Test Process

The two evaluators (Kujala and Härkönen) both apart from each other read the same three MindFit Analyses and filled the Questionnaire Form per each analysis. This version of the questionnaire contained limited explanations for the ten Criteria in question (see attachment 3). Based on the results it was quite clear that the results looked rather consistent. Because there were only three analyses and two evaluators, no statistical test was possible to be run. The minimal differences in the results could just be seen by investigating the data manually. Even though the results indicated that the study could move on to an actual official reliability test (with 4 evaluators and 20 analyses), it was decided to clarify the instructions of the form. This clarification contained the detailed Criteria Descriptions which should not leave much space for misunderstanding about what the Criteria should measure and how each of the competences in the Criteria should be rated (Level Descriptions per Criteria) on a 7-step scale.

Based on Pre-Reliability test the major change to Questionnaire form was to define separate descriptions of the score levels per each Criteria. It was thought to be essential that the evaluators understand what each score exactly means for the particular Criteria in question. For example, in the Pre-Reliability test Questionnaire version the score 3 was the same for all of the first five competences: "*Has an overall level understanding of the issue. Can perform the task rather independently if getting support when needed*". After modifying the explanations in order to make the actual Reliability Test version of the Questionnaire, the score descriptions were more specific. In the final Questionnaire form the score 3 for the Communication skills Criteria implies '*Can take minor responsibilities regarding company's information sharing*' and the score 3 for the Management skills implies '*Can independently lead small team in his/her own competence area*'. The process of developing the Questionnaire form can be seen by comparing the Questionnaire form versions, i.e. by comparing Pre-Reliability Test (Attachment 3) with the final Reliability Test (Attachment 2).

4.3 Choosing the Statistical tests

To measure if the questionnaire contains test items (such as Criteria) that correlate with each other, the internal consistency can be studied to estimate the reliability (Nunnally & Bernstein, 1994). To be noted that if the test items assess different attributes, the internal consistency between items is poor, indicating that the items don't follow a common theme (Cook, 2004). However, by measuring consistency we can find out how the measurement tool, the questionnaire in this case, helps us when searching differences between raters in a situation where there should be differences. And usually, there are differences since evaluations are subjective rates. The internal consistency between raters can be measured using inter-rater reliability estimation (IBM SPSS Statistics, 2011), such as intraclass correlation coefficient (ICC) that is based on Analysis of Variance (ANOVA) model (Nichols, D., 1998).

Statistical tests can be divided into parametric and non-parametric ones. According to Nummenmaa (2004) parametric tests are more powerful than non-parametric ones, so those should be used if the data has been collected in a form that fills the expectations for the parametric tests. In order to get as precise results as possible while comparing the data that humans have rated, it should be used parametric tests always when it is possible. In order to use parametric test, the base routine is to test if the data follows normal distribution. In our case, the Kolmogorov-Smirnov test didn't give support to the normality of data (0,253; Sig. = 0,000). However, since it was not found a suitable non-parametric test corresponding the ICC, it was decided to use parametric ICC. With ICC it is possible to investigate qualitative material quantitatively, as soon as the data is operationalized properly (Coolican, 2005). Another factor affecting proper statistical test is the form of data, whether it is continuous or discrete. As described above (Choosing the Rating Scale) the 7-step scale was taken into use since it is closer to continuous variable than the base 5-step discrete Lickert for example.

4.3.1 Intraclass Correlation Coefficient Test

To measure inter-rater reliability the Intraclass Correlation Coefficient (ICC) test was chosen. Although the actual data is not completely continuous and does not completely distribute normally, it is the common way to measure the reliability of a questionnaire of several scale values. For example, according to Niitamo (2003), the Cronbach Alpha is to be used e.g. in multi-steps Likert-scale. To be noted that in SPSS the Cronbach Alpha works parallel with the ICC (i.e. it gave exactly the same values as ICC).

The ICC also seemed to work as a suitable tool to measure evaluators' consistency, since it considers the possible measurement error related to rating scale usage of the rater. The *generalizability theory*, from which intraclass correla-

tion follows, explains the interpretation principals regarding the rating scales. For example, if some rater consistently uses the scale below or above the average scores compared to other raters' scores, due to own understanding of the Criteria or due to score level descriptions, the ICC takes this behavior into a count in a calculation model. Thus the way how the rater uses the scale affects beside the score itself. These kind of differences are difficult or even impossible to observe manually and the test may still be reliable even if the rates differ in scores (Nunnally & Bernstein, 1994.).

The data was expected to be rated by random evaluators who each rated all the same analyses. Also the possible systematic difference between scores was not considered to be in present. Thus, the Two-Way Random model with the Consistency option was chosen in SPSS, see Nichols, D. (1998) and Shrout, P. & Fleiss, J. (1979).

4.3.2 Factor Analysis

In order to see if the rates form any kind of sub groups, such as division related to practical and personal competences, the factor analysis was done from the mean values of the raters per each competence of the Criteria. It was investigated such dependences between items, that couldn't be seen just by manual observation of the data. However, the diversity between Criteria types (competences) was drafted to be different in purpose, in order to evaluate different types of competences and qualities of a person, and thus it was not expected to be found such a division between practical and personal competences in Criteria. In fact, while interpreting the factory analysis report, no such division was found. This supports the diversity idea of the Criteria, i.e. supports the idea of measuring several different types of competences of a person.

4.4 Evaluators

Evaluators who rated the analyses using the questionnaire form present a random sample of population, i.e. an employer's personnel who are responsible for recruiting professionals into their organization. Three of the four evaluators didn't know who they were evaluating for. Their evaluation was purely based on the text in a MindFit Analysis. On the other hand, one of the evaluators (undersigned) was the same person who wrote all the analyses, so he had met all the candidates in person. Still, the purpose was that he would fill in the questionnaire just by interpreting the analysis text, not by memorizing any candidate's skills and behavior observed while meeting them. Major of the twenty analyses scored were written that long time ago (over a year, and some of them over five years ago) that it was also impossible to remember the behavior of the analyzed candidate well.

The evaluators have following background. The undersigned has completed an IT Engineer degree and studied psychology and cognitive sciences. He has over 15 years' experience from the IT field, e.g. as an engineer, project manager, business development manager, and recruitment/personnel evaluation analyst. One evaluator has studied education sciences and worked two years in HR business and as a personnel analyst plus a few years as a coordinator at the university of applied sciences. The third evaluator came into HR business from the psychology department of the university and has about four years of experience from various HR tasks, including personnel analyzing. The fourth evaluator has a commercial Master's degree and several years of work experience from the HR field and marketing, mainly from assistancy and coordinative responsibilities, but she also has experience of pre-selection of job candidates. She was the only evaluator who had no experience from working as a personnel analyst or writing personnel analyses.

4.5 Implementing the reliability procedure

Totally, the same twenty (20) analyses were rated by every four (4) evaluators. The evaluators filled in a questionnaire form per each of the twenty analyses written about job candidates. The evaluators rated analyses independently, getting the same instructions and they were not able to discuss with each other while filling in the forms. They filled in the forms remotely at home.

4.6 Testing the hypothesis

According to the hypothesis, the first five Criteria was estimated as easier to understand, i.e. easier to be rated by evaluators compared to the last five. Thus better consistency scores were expected from the first five Criteria. This assumption was based on the type of Criteria, since the first five seemed more *practical* oriented competences (such as Core skills, Technical skills and Management skills) compared to the more *personal* oriented Criteria in the last five (such as Personal qualities and Orientation). Thus the first five practical Criteria was thought to be reported more clearly in an analysis report. It was also assumed that the practical Criteria were easier to be clarified in an interview and testing session than the personal Criteria.

5 RESULTS AND DISCUSSION

The reliability results of the study are very promising giving almost every competence of the Criteria over 0,70 intracorrelation coefficient that implies the consistency between the evaluators' scores. There can also be seen some differences between practical and personal competences supporting the hypothesis of practical competences to be more reliable.

5.1 Reliability test

The reliability test was based on measuring inter-rater reliability using intraclass correlation coefficient function in SPSS. The results show over 0,70 reliability to eight of the ten competences in the Criteria. If the results are to be interpreted with two decimals, there would be nine competences reaching the limit of 0,70. The competence number 7 (Orientation) was the only competence that got rather low result (0,365), since the second poorest competence number 6 (Personal qualities) was close to 0,70 having a value of 0,695 with three decimals. The results are presented in the table 1.

Reliability Study: Inter rater reliability - Results

SPSS Method: Analyze-Scale-Reliability Analysis-Statistics-Intraclass Correlation Coefficient-Two-Way Random-Absolute Agreement / **Consistency**

	Reliability Study - Results (Ability, skill, trait, appl.)	Intraclass Correlation Coefficient (Absolute Agreement)	Intraclass Correlation Coefficient (Consistency)
1	Core skill	0,732	0,724
2	Technical skills	0,816	0,847
3	Communication skills	0,839	0,845
4	Presentation skills	0,899	0,908
5	Management skills	0,857	0,889
6	Personal qualities	0,666	0,695
7	Orientation	0,324	0,365
8	Career expectations	0,662	0,74
9	Risks (0 = high; 6 = no risk)	0,671	0,801
10	Applicability	0,814	0,825

Table 1 The reliability results between raters using intraclass correlation coefficient.

As described in Methods, the Consistency function using Two-Way Random method was used in SPSS (ICC values in black letters). Just for the comparison, the test was run also with Absolute Agreement function (grey values).

5.2 Interpretation

The results related to the consistency of the raters' scores are depending on how clearly the analysis text describes the each competence in the Criteria. The more in detail the analyst has managed to clarify the skills and competences of the candidate, the more clearly the analyst has been able to document the competence in an analysis. The scores in the reliability test seem consistent in general, although some differences between the scores can be seen. There is one clear exception (Orientation), which is clearly below the acceptable reliability limit. Below each of the competences in the Criteria has been interpreted based on the reliability value they got from the consistency rates:

1. Core skill, reliability value 0,724
   One of the most relevant competence of the personnel analysis is the core skill which explains the skill level of the candidate in the area most famil-

iar to him or her. The core skill describes what kind of work results an employer can expect from the job candidate in his or her core expertise area. The core expertise has been tested using several methods during the personnel evaluation (e.g. competence inventory, in-depth interview related to core field and the presentation test). However, since the whole session lasts maximum of one and half an hour, it sets some limits for getting the complete understanding of the candidate's skill level resulting ocassionally some inaccuracy in the analysis.

2. Technical skills, reliability value 0,847
   While evaluating competences in a wider perspective, the testing and interview session seems to provide analysis text that can be understood very similarly by the raters what comes to description of the technical competences of the candidate.

3. Communication skills, reliability value 0,845
   The ability to communicate can rather well be seen in an intensive interview session that covers both the formal structured questioning and free discussion. Thus the outcome has also been reported clearly in an analysis, providing that the communication part has been understood very similarly amongst the raters.

4. Presentation skills, reliability value 0,908
   The extremely high reliability value related to the presentation capabilities of the candidate could be due to fact that the presentation test gives a physical result (picture). In the test the candidate presents e.g. some IT architecture by explaining its functions and using pictures. This can be interpreted rather objectively, especially if the evaluator has substance knowledge in the field.

5. Management skills, reliability value 0,889
   Contradicting to Wohlers & London (1989) who suggests that the leadership competence gains poor consistency amongst raters, the management competence provided excellent consistency value in our rating. The poor results in Wohlers' & London's study could be related to many things, such as how the raters have understood the management as an activity at work (if they have understood it similarly), or that the leadership is, in fact, a broader issue to understand than pure management, which can be seen as a more practical operation.

6. Personal qualities, reliability value 0,695
   Although close to the acceptable reliability limit of 0,70, the results show some improvement needs when it comes to clarifying a person's qualities as a worker. The low result may be due the fact that the evaluation of personal qualities is not structured clearly into sub-categories in a per-

sonnel evaluation session (such as learning skills, personal interests, flexibility etc.).

7. Orientation, reliability value 0,365
   Orientation was the only competence in the Criteria clearly below the acceptable level. The person's orientation has obviously not been stated clearly enough in an analysis. The explanation could be quite clear, since the orientation has not been handled thoroughly in a personnel evaluation session. In order to report the person's orientation clearly in an analysis text it would need to be discussed intensively in an interview by clarifying the true interests and goals of the job candidate. Such an intensive clarification would of course require more time. It is also a kind of information that could be difficult or even impossible to gain reliably using questionnairies.

8. Career expectations, reliability value 0,74
   This quality is quite close to Orientation as a personal 'competence', but has limited to more narrow area which makes the evaluation easier. Thus it has been reported understandably in an analysis providing rather consistent values amongst the evaluators. The career related interest is one of the most important issues regarding the long term work relationship.

9. Risks, reliability value 0,801
   The technique of discovering possible risks related to the job candidate has not been structured clearly in the personnel evaluation session. Still, it obviously has been clarified well in an analysis due the good result in the consistency scores. The clearly written analysis and the good result may be related to the free dialogue between the interviewer and the job candidate. In a less formal discussion the true interests or the candidate's past risky behavior may come up, and can thus be reported very clearly in an analysis.

10. Applicability, reliability value 0,825
    Probably the most relevant information for the employer to know about the job candidate is his or her applicability for the job and the industry field in general. Almost all testing methods and interview techniques provide important input for the evaluator in order to report the general applicability of the candidate. Thus the result expresses a very good consistency between the readers of the analyses. To be noted that the applicability is much more than experience from the field. In fact, in case of a young candidate, the applicability could be high without any relevant work experience, e.g. if the candidate has indicated other essential abilities (such as an ability to use interdisciplinary knowledge) and motivation regarding his or her suitability for the field.

Although the consistency results are quite potential and support the similar approach to be used in future as well, there are certain observations that have to be considered seriously in order to improve methods to be more accurate. The practical oriented competences all gained consistency values over 0,70, but personal values, *Orientation* and *Personal qualities* were left under it, 0,365 and 0,695 respectively, and are thus to be taken under closer investigation.

5.3 Hypothesis

Based on ICC results, the hypothesis is partly supported since all the practical competences (1 to 5) seen in Table 1 reach the ICC over 0,70, but the competences defined as personal contain two values below the reliability limit of 0,70; Personal qualities (0,695) and Orientation (0,365) are left below the limit. Otherwise there are rather high values in personal qualities as well, such as Applicability (0,825) and Risks (0,801). Only the consistency rate for the Orientation is a clear exception within the Criteria, which is rather consistent as whole without a clear difference between the practical and personal division.

6 CONCLUSIONS AND FUTURE DEVELOPMENT

Competence based personnel evaluation is an important decision making tool for the organizations and supervisors hiring new professionals and managers. Personnel analyses are reports written about job candidates based on an interview and evaluation session. In the study of investigating the reliability of the personnel analyses based on MindFit Consulting Ltd personnel evaluation methods, the purpose was to clarify, if the different persons who read the analyses interpret them similarly. This is the key issue to know since the personnel who read the analyses in practice, are supervisors in the customer companies hiring personnel partly based on the content of the analyses.

In order to measure the reliability, the questionnaire form was built to measure how different readers rated the content of the analysis. Since the analysis is written in a verbal mode without numeric values of the job candidate's skills or competences (except for the work experience in years), the questionnaire form had to be operationalized for the scoring purposes. The evaluators had to be able to give scores for the different competences picked up from the analysis. The operationalization produced ten competences that can be seen as practical competences and personal competences. However, the evaluators were not informed about the different nature of the competences. The practical competences formed the first five of the Criteria (Core skill, Technical skills, Communication skills, Presentation skills and Management skills) and the personal competences presented the last five Criteria in the questionnaire form (Personal qualities, Orientation, Career expectations, Risks and Applicability).

The measurement is based on the consistency between the scores of the four evaluators who all read the same twenty (20) MindFit Analyses. The inter-rater reliability using ICC was set to a limit of 0,70 based on other reliability studies regarding questionnairies. Although Nunnally & Bernstein (1994) refers to the reliability level of 0,70, there are other limits mentioned in the literature as well. For example Coolican (2005) refers values 0,75 to 1 for the correlation coefficients. The evaluators scored the analyses in the scale of seven options. The

meaning of each scale changes gradually from 0 to 7 (from 'no skills' to 'expert'). The scale size was expected to be suitable for getting rather detailed scores, still remaining in limits that a person can keep in mind while reading scale instructions and scoring the analysis.

The main result of the study is that the Criteria with ten different competences scored by raters are mostly reliable. This means that persons who read the analyses rated the competences similarly and thus it can be assumed that they understood the content of the analyses similarly. More specifically, based on Inter-rater reliability test using ICC, all the practical competences (first five) in the Criteria reached the proper reliability limit set to 0,70 having values from 0,724 (Core Competence) to 0,908 (Presentation skill). Also, the majority of the personal competences (last five) got very reliable results, with only one of them (Orientation) getting a value below to the common reliability limit of 0,70, with a value of 0,365. The Person description competence was left only slightly below the limit with the value of 0,695 (also this would reach the limit of 0,70 if counted in two decimals). The result supports the overall reliability and is in line with the hypothesis about practical competences being more reliable, although with minor exceptions.

6.1 Practical and personal competences - hypothesis

The expectation and hypothesis were that practical oriented competences are easier to be evaluated compared to more personal oriented competences, and that those provide better results in the reliability test. As noticed in Wohlers & London (1989), it is not obvious, that those competences that feel easy to be evaluated will automatically be rated consistently amongst raters. However, in our study, competences that were classified as practical oriented were expected to be documented in a more vague way in an Analysis as well. The hypothesis was based on the analyst's own experiences gathered from the evaluation sessions during the past ten years, as well as on literature.

The literature regarding the rating of skills support the base idea of the hypothesis, although there were no exactly the same competences rated in other studies. For example, Viswesvaran et al. (1996) gave somewhat supportive observations by reporting that communication and interpersonal competences were less reliable than productivity and quality. Although in MindFit Criteria the *Communication* competence is classified as practical (and thus more reliable) competence compared to personal ones, the idea that productivity and quality are easier to be measured based on results and numbers, will generally support the base idea that practically observable skills can be assessed more reliably than (less practical) personal qualities of the person. Wohlers & London (1989) supports perhaps even more clearly the hypothesis by suggesting the exact same idea (with some exceptions) that characteristics that are easy to rate provide

more consistency amongst raters than those that are difficult to rate. Although the competences to be rated varied somewhat from the competences in MindFit Criteria, one of the remarkable results in Wohlers' and London's study was that if the characteristic or behavior is observable by the rater (e.g. written skills, oral presentation and energy) it is also rated more consistently within raters.

So, there were no the exact same competences rated in other studies, and some of the competences classified as practical ones in MindFit were handled as more personal or psychological ones in other studies. The general outcome from these studies, however, supports the idea that observable competences are easier to be evaluated since those can be rated by observing the person's behavior, compared to more psychological competences. In the MindFit study, the very observable practical competences measured the way described in *Evaluation Methods*, were *Presentation skills* and *Communication skills*. Those could be directly observed in an intensive interview and presentation test session. Moreover, the *Core skills* and *Technical skills* are also competences that can be observed during MindFit's interactive presentation session, with only the time (5 to 10 minutes) limiting the amount of data gained about a candidate's skills. However, the *Management skills* is the only practical competence that cannot be observed directly using MindFit evaluation methods. Still, Management skills got very consistent value (0,889) in the reliability test suggesting that the Analyst has described the skill very understandably - the question is, would that description correspond to the reality state of the candidate's management skills.

Since the personal skills are more and more appreciated when hiring new personnel into today's industry, such competences could be measured beside the practical ones. The contextual performance (e.g. Niitamo, 2003 and Motowidlo et al., 1997) for example is seen as one major factor that should be observed. Compared to personal Criteria in MindFit, the *Personal qualities* and *Applicability* competences cover the meaning of the contextual performance quite well, and thus they seem relevant to be included in the Criteria. The other personal competences (*Orientation*, *Career expectations* and *Risks*) that MindFit used in this study support and specify this personal Criteria further. They seem to be competences that are not easy to be evaluated by observing the candidate's behavior. Thus they are in line with Viswesvaran et al. (1996) and Wohlers & London (1989) and can be thought as personal or psychological competences, as the division has been defined in the MindFit study.

The results of the reliability test support the hypothesis for the major part of the Criteria. The most consistent value, *Presentation skills* (reliability value 0,908) belongs to the category of practical competences and the absolutely most unconsistent result got for the *Orientation* (reliability value 0,365) to the personal competences category. Four of the five practical competences gained better values than any personal competence. Only *Core skills* from the practical competences was left below a few personal competences, but still got rather reliable

consistency value of 0,724. Alltogether, the hypothesis was in line with the literature on the subject and with the results in this study, although it was not completely supported.

6.1.1 Silent knowledge

Silent knowledge, or perhaps the better known term from the psychology, the tacit knowledge, is knowledge that someone has but can not easily spread out to others. As presented earlier, the more the expert level job task in question, the more the person needs to hold tacit knowledge (Eraut, 2000). Thus one of the key issues to be evaluated in person is the tacit knowledge. In order to get as realistic understanding as possible about the candidate's true knowledge base and potential, the evaluation methods should reveal how the person shares his or her knowledge in a formal situation (such as in a formal interview, in front of an audience, etc.), compared to his or her beharvior in a less formal situation, such as in a free conversation, answering questions or presenting something (e.g. describing something using drawings) – all methods that belong to the evaluation repertoire used in this study, e.g. in a form of presentation test that also simulates a typical work-scenario while the interviewer asks questions and comments the scenario or picture that is under presentation by the candidate.

A common finding e.g. in Khumalo (2012) and Wagner and Sternberg (1985) is that the tacit knowledge is related to practical skills and behavior oriented competences more than it is to personal ones. According to Eraut (2000), it can also be observed better in practical behavior than in personal qualities. Since the practical competences constitute an important role for the personnel evaluation, the tacit knowledge can bring significant added value for the evaluation.

6.2 Discussion

The study provides an important information about the content of MindFit analysis. It clarifies which areas people understand most consistently, as well as points out the areas where people have challenges to get common understanding. Fortunately, the most important parts of the Criteria, the practical competences, were all understood consistently enough (consistency values between 0,724 and 0,908). The areas requiring more investigation and probably some improvements as well were related to personal qualities, especially to job candidates' Orientation. The results give an important input about potentiality of the methods as a recruitment tool, as well as versus to the more traditional methods. However, some issues related to methods of the study process as well as to methods of making analyses in general have to be considered while interpreting the results.

6.2.1 The relevance of the evaluators

The sample of the evaluators in this study should represent people who read personnel analyses in practise on the customer side. Customer's people reading the analyses are either technical managers/supervisors of the employees or companies' HR personnel, or both. The common thing for all the customers' personnel is that they are influencing or responsible for the recruitment, where personnel analysis works as a tool for making the decision of hiring a person. Two of the evaluators in the study has background with recruiting people. One evaluator only has real IT expertise and practical experience of it, while the others are trained to the core technical areas theoretically. All four of them has professional experience from HR work in general. This is somewhat in line with the background of the people who usually read analyses, although not completely. What is really different between evaluators and customer personnel is the experience of writing personnel analyses. Three of the four evaluators have experience of writing personnel analyses while probably not one of the customers' personnel has such an experience. It can be questioned if the experience of writing analyses influenced to their ability to interprete analyses. Although the consistency results between the four evaluators didn't differ that much, it could still be investigated possible differences between the one evaluator with no experience of analyzing and the experienced analysts. Moreover, to get more reliable results, a few more evaluators with diverse background could rate the same analyses.

6.2.2 MindFit personnel analysis as a recruitment tool

At this point, there are no statistics about the real validity of the analysis versus the employee's real work performance. However, this study is one step forward in understanding the real validity of analyses. To be noted that the total validity i.e. ecological validity is different from the validity of the Questionnaire form described here. Even if the form gets high validity value, the MindFit analysis and the Questionnaire form are not valid if the Criteria in the analysis and the form are not related to the competences required at work. In order to be a really meaningful recruitment tool the analysis has to cover competences that are relevant for the employer to know about the job candidate.

This study, however, presents, that the main object of the investigation, the MindFit Analysis, is understood very similarly amongst different readers (i.e. evaluators who scored the analysis text on a seven-step scale). Although some personal competences were left below the expected reliability limit, the most important competences from the practical work task oriented perspective got high values in the consistency analysis. This result is a good base to start with when investigating the actual validity of the analyses later.

6.2.3 Traditional evaluation methods vs. methods used for analyses

The traditional evaluation methods are focused on to theoretical knowledge of psychology and personnel evaluation, as well as to the analyst's work experience as a personnel evaluator. Not even the common competence based evaluation methods mention the work field or industry related knowledge as a requirement for the personnel analyst (Honkanen, 2005). Conversely, beside the theoretical knowledge related to psychology, cognitive science and psychometrics, the MindFit Analyses are written based on the analyst's knowledge and experience from exact the same industry field where the evaluated candidates are specialized. This extra dimension may give the analyst the opportunity to cover the knowledge base of the candidate more in-depth in comparison to situation where the evaluation is done without the specific field knowledge. This superiority has not, however, been confirmed yet by *validating* the analyses in practise. Still, as e.g. May et al. (2014) describes, the substance knowledge of the analyst helps in getting more consistent scores about behavioral function between raters (*reliability*).

6.3 Future development

Here are some ideas and discussion about how the personnel evaluation and analyzing methods could be developed in future, as well as how the validity studies could be get into a more specific level.

6.3.1 Improvements to analysis methods

Although every practical competence in the reliability study gained consistency value over 0,70 which can be seen as satisfactory level for the factor, the *Core skills* competence could do a little better (now 0,724). Core skills of the job candidate are more important the more expertise level tasks are in question. One way to gain better results could be to reserve more time for an analyzing and testing session to concentrate finding out the candidate's core knowledge. For example, some more time during the presentation test, spent e.g. specifying the questions more deeply and mainly going through the knowledge areas that belong to a core expertise of the candidate, could give a clearer picture for the analyst about the candidate. That could be done without increasing the total time of the testing and interview session by priorizing the time sharing more effectively. To be noted that if more time is spent in an interactive presentation session, it may not decrease the total outcome since the presentation situation provides valuable information about other competences, such as communication, as well.

However, the major concern raised from the reliability study is related to personal competences, such as *Orientation* (consistency value 0,365) and *Personal*

qualities (consistency value 0,695). It can also be questioned, if those should be included at all into the analysis, or if they should be replaced with more suitable competences, that for example, would support better contextual performance and/or organizational citizenship that were appreaciated in the international studies mentioned earlier (Niitamo, 2003 ; Motowidlo et al., 1997 & Mark Cook, 2004).

6.3.2 Improving the relevance of the Criteria

In this study only the reliability of Criteria in the Questionnaire form was studied (using ICC), i.e. how consistently evaluators understood the Criteria. However, it can be questioned what the effect would be if the relevance of each competence in the Criteria were defined together with customers. In order to clarify if the Criteria measures issues that are relevant for the evaluation process and recruitment, the information can be gathered in the second part of the form, see figure 5 (the whole questionnaire form is presented in an Attachment 2). In the form the supervisor of the employee (customer) is asked how important the specific competence in a Criteria is while evaluating the particular candidate – note the question in the form: *Relevance of criteria at the work assignment?* This relevance is one of the factors affecting the total *validity* of the analysis. Note that this part of the Questionnaire form (Table 2 in figure 5) was not filled in by evaluators who rated the analyses in this study, which has the purpose to clarify the *reliability* of the Questionnaire form *Criteria*. The Table 2 can be filled in by supervisors in the next phase while investigating the *validity* of the Analyses (In Dr. Thesis).

	Table 1								Table 2		
	Criteria	The level of criteria (ability, skill or trait)							Relevance of criteria at the work assignment		
	(Ability, skill, trait, appl.)	0	1	2	3	4	5	6	Very important	Good to know	Not relevant
1	Core skill						x				

Figure 5 The Relevance of criteria, supervisor marks one of the three options per criteria.

There is also literature about other competences that can be analyzed in person. For example, instead of analyzing Personal qualities as defined in the questionnaire form (Attachment 2/2), Eraut (2000) presents *personal knowledge* as a competence that is defined as a cognitive resource which a person brings to a situation that enables them to think and perform. The definitions of MindFit Personal qualities and Eraut's Personal knowledge are, however, somewhat similar since both are related to adaptability of work practises. Still, to develop analyses further, the content should be clarified together with several customers, and a criteria based on those definitions should be defined.

6.3.3 Advanced competence inventory

The current competence inventory does not ask a candidate's opinion about his or her skills, although it has been pointed out (also in this study) that the years of experience does not measure reliably the level of skills on the competence in question. The simple and valuable improvement would be to gather such information in a competence inventory phase while the candidate leaves his or her CV and application to MindFit. Such a competence list that collects candidates' own estimations about their skill levels is already in use in MindFit's own CV template, see figure 6.

Competence	Work exp.(years)	Level (1-5)	Industry field	Last in use
Header A (e.g. Web)				
<Competence 1>	<e.g. 5>	<e.g. 3>	<e.g. IT Security>	<e.g.2012>
<Competence 2> ... **Header B (e.g. dBs)** <Competence 1>				
<Competence 2>				
...				

Skill levels (Level 1 – 5):

1. **Junior,** *ready to work, but little to none work experience; work experience less than 1 year.*
2. **Professional**, *does not possess the experience to fully take advantage of the tool/environment/skill in question; work experience 1-3 years.*
3. **Senior**, *can implement and develop with the tool/environment in question. Can guide others and is able to carry out singular tasks independently; work experience 3-5 years.*
4. **Spesialist,** *is expert in particular tool/environment. Ready for independent and responsible working; work experience over 6 years.*
5. **Guru,** *is ready for demanding tasks. Can help and guide others in the most problematic cases. Exceptional understanding of particular tool/environment and able to carry out very responsible tasks.*

Figure 6 Competence list with skill level (Level 1-5) estimation.

6.3.4 Differences between IT professionals

Although some practical competences may be rather easy to observe and mesure, as the example presented in *Competence Measurement* referred to items manufactured per day or sales per week, the whole picture may not be that clear just based on performance results. For example, if programmers are rated based on the amount of code they produce per day (practical competence), the scoring could leave the quality or validity factor out, i.e. the questions like: is the amount of code made with such a quality that is useful for the product and the organization, or, does it provide a solution for the problem in question, or, does it require adjustment at later stage, and so forth, should be considered.

In this study it was referred to some old studies about the differences between programmers' performance (Sackman et al., 1968 ; Schmidt et al., 1980 ; Dickey, 1981). Today's programmers use a lot higher level languages although those used in 60's are to be used today as well, especially in a low level, close to hardware interface. The more recents studies, such as Witt & Burke (2002) should be investigated to get a more realistic picture about differences between personnel in the industry of our times.

6.3.5 The predictive value of personnel analysis

One of the key information usually wanted from personnel analyses is to get an understanding about the candidate's future potential as an employee. This gets paradoxical if it is to believe the Finnish Psychological Association's (FPA) recommendations not to make complete estimates about the candidate's future behavior based on competence evaluation methods. Elsewhere Stanovich (2004, 189) refers to Dawes (1994) saying that not one evaluator (psychologist or not) should make confident predictions about the future. From that perspective the personnel analysis should only report about the candidate's current behavior and potential, as well as refer to the statistics that can be proved e.g. by the person's work history data. Also, the statistics (Dawes, 2005) support the idea that clinical observation that is based on personal oriented evaluation can not beat the power of more practical competence oriented actuarial prediction. Thus the practical competence based evaluation can be expected to predict the candidate's future performance more accurately than personal qualities oriented evaluation methods.

However, as presented in this study, there are important personal qualities in people that are useful for the organization. Contextual performance was an important quality appreciated by many researchers (Niitamo, 2003 ; Motowidlo et al., 1997 ; Cook, 2004). Still, in order to gain the real predictive value of such performance, it should be divided to sub competences or personal features (such as conscientiousness and agreeableness) that would need to be measured reliably. *Curiosity* was also presented as a personal quality that would predict many important competences useful at work (Mussel, 2012). So, there are, of course, important personal qualities that are for the employer to know while recruiting new employees. The challenge for the future method development is how those qualities could be measured reliably. This rises a relevant question: Should more time be used to measure and analyze practical oriented competences to get more useful results for the recruitment?

On the other hand, there are studies (May et al., 2014) supporting such evaluation principal, that the analyst is experienced on the work field and tasks related to the job in question. Thus, while evaluating job candidates, it could be a strong benefit for a personnel analyst with subject field knowledge (e.g. IT) compared to one with no specific field experience. Still, for some reason, the professional field experience is not considered in the FPA recommendations. The future studies could be set to define the ideal combination of field knowledge (such as IT technology) and psychological knowledge for the analyst in order to make more reliable and valid estimations about job candidates' future work performance.

To get more practical support for the concept of the future estimations related to work performance, the study about the validity of personnel analyses should

be executed. Such a study should investigate how the analysis of the job candidate correlates with the real work performance of this person as an employee. Cook (2004) points out that the criterion validity predicts the productivity. The criterion validity can be divided in three types; predictive validity, concurrent validity and retrospective validity. Especially the predictive validity seems convincing since it is about how the subject to be tested manages in practise when measured later, e.g. at work after a few months (Ackerman & Beier, 2006). The test data about the subject in concurrent validity test is collected at the current moment and the practical performance of the subject is measured at the same time (e.g. concurrently), while the retrospective validity investigates the history data comparing it to the current performance (Cook, 2004.). If measured using all three types of criterion validity tests, it can be compared how valuable those different methods are. The retrospective and concurrent validities are, however, quicker and probably a lot more cheaper to produce since the predictive validity test requires time for the measurement to be done at later stage.

6.3.6 Future studies and reading

Michael Eraut writes about one of the most important issues in personnel evaluation, the *tacit knowledge*. For example, Eraut's article used in this study handles also the rapid action and its relation to tacit knowledge (Eraut, 2000). The tacit knowledge is in general the area that should be investigated more in-depth, especially related to methods of how to recognize one in a person.

As Leong et al. (2013) suggested, *orientation* of the person could tell much about that persons career choices. It is thus the area that should be studied more thoroughly. Their tool for clarifying such orientation as well as their results were promising related to reliability of such a study, although they didn't test the concept related to real choices in work life. It would be interesting to know how much the orientation and career interests of the job candidate really affect the work performance, motivation and the long term commitment between the employee and the employer.

The literature presents also other competences that are not investigated in this study and which could be valuable predictors for work performance. There are studies that have provided optimistic results about general mental ability (GMA) test predicting work performance reliably (Schmidt & Hunter, 2004 ; Witt & Burke, 2002). Also, according to Schmidt & Hunter (2004), conscientiousness is the most effective personal quality of five-factor model (FFM) to predict work performance. Thus, beside the already mentioned competences (in chapter 2.2.2), GMA and conscientiousness could be taken into the criteria of personnel evaluation as well, if further studies support their effectivity to predict work performance. It may require a proper psychological and/or cognitive oriented test to measure those, if pure competence based tests cannot cover

such qualities. Thus, further studies about the ideal relationship between the competence based evaluation and psychological tests should be proceeded.

7 REFERENCES

Ackerman, P. & Beier, M. (2006). Methods for Studying the Structure of Expertise: Psychometric Approach. In Ericsson, K., Charness, N., Feltovich, P. & Hoffman, R., *The Cambridge handbook of Expertise and Expert Performance*, 147-165. New York: Cambridge University Press.

Bretz, Jr., R & Judge, T. (1998). Realistic Job Previews: A Test of the Adverse Self-Selection Hypothesis, *Journal of Applied Psychology, 83* (2), 330-337.

Cook, M. (2004). Personnel Selection. Adding value to people.
Old and new selection methods, 1, 10.
Test of mental ability, 99.
Criteria of productivity, 228. Chichester: John Wiley & Sons Ltd.

Coolican, H. (2005). Research Methods and Statistics in PSYCHOLOGY. *Observational methods – watching and being with people,* 114. *Questionnairies – asking people using psychological scales,* 171. *Analyzing qualitative data,* 468. London: Hodder & Stoughton Educational.

Cowan, N. (2000). The magical number 4 in short-term memory : A reconsideration of mental storage capacity. *Behavioral and brain sciences, 24,* 87-185.

Cummings, J., Teng, B (2003). Transferring R&D knowledge: the key factors affecting knowledge transfer success. *J. Eng. Technol. Manage, 20,* 39-68.

Dalessio, A. (1994). Predicting insurance agent turnover using a video-based situational judgment test. *Journal of business and psychology, 9* (1), 23-32.

Dawes, R. (1994). House of cards : *Psychology and psychotherapy built on myth.* New York: Free Press.

Dawes, R. (2005). The Ethical Implications of Paul Meehl's Work on Comparing Clinical Versus Actuarial Prediction Methods : *Journal of Clinical Psychology, 60* (10), 1245-1255.

Dickey, T. (1981). Programmer Variability. *Proceedings of the IEEE, 69* (7), 844-845.

Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology, 70*, 113–136.

Ericsson, K. (2006). The Influence of Experience and Deliberate Practise on the Development or Superior Expert Performance. In Ericsson, K., Charness, N., Feltovich, P. & Hoffman, R., *The Cambridge handbook of Expertise and Expert Performance*, 683-703. New York: Cambridge University Press.

Horn, J. & Masunaga, H. (2006). A Merging Theory of Expertise and Intelligence. In Ericsson, K., Charness, N., Feltovich, P. & Hoffman, R., *The Cambridge handbook of Expertise and Expert Performance*, 587-611. New York: Cambridge University Press.

Hakkarainen, K., Lonka, K. & Lipponen, L. (1999). Tutkiva oppinen. *Älykkään toiminnan rajat ja niiden ylittäminen,* 24. (1.-3. painos). Porvoo: WSOY.

Honkanen, H. (2005). Henkilöarviointi työelämässä. *Henkilörviointimenetelmät,* 99, *Liite 2. Arviointitavat ja niihin vaadittava pätevyys.* Helsinki: Edita

Honkanen, H. & Nyman, K. (2001). Hyvän henkilöarvioinnin käsikirja. *Kuka on pätevä tekemään arviointeja,* 77-81. *Arviointimenetelmät ja niiden luotettavuus,* 90. Helsinki: Psykologien Kustannus Oy

IBM SPSS Statistics, (2011). *Measures of reliability in scale problems, Release 20.0.0.*

Khumalo, F. (2012). The Tacit knowledge transfer process : How Middle Managers Facilitate Knowledge Sharing in Organizations. *A Dissertation Presented in Partial Fulfillment Of the Requirements for the Degree Doctor of Philosophy*. Capella University: UMI Number: 3512454, ProQuest LLC

Leong, F., Pearce, M. & Huang, J. (2013). Assessing Scientist and Practitioner Orientations in Industrial/Organizational Psychology. *Journal of Career Assessment, 21* (3), 452-468.

May, M., Sheng, Y., Chitiyo, M., Brandt, R. & Howe, A. (2014). Internal Consistency and Inter-Rater Reliability of the Questions About Behavioral Function (QABF) Rating Scale When Used by Teachers and Paraprofessionals. *Education and treatment of children, 37* (2), 347-365.

Motowidlo, S. J., Borman, W. C. & Schmit, M. J. (1997). A Theory of Individual Differences in Task and Contextual Performance. *Human Performance, 10* (2), 71-83.

Mussel, P. (2012). Introducing the construct curiosity for predicting job performance. *Journal of Organizational Behavior, J. Organiz. Behav., 34*, 453-472.

Nichols, D. (1998). Choosing an intraclass correlation coefficient. *SPSS Inc.*, Nr 9.

Niitamo, P. (2003). Henkilöarviomenetelmät työelämässä.
Työmenestymisen ennustaminen, 17.
Hiljainen tieto, 100-101.
Testimenetelmien luotettavuus, 124. (1. painos).
Jyväskylä: Gummerus Kirjapaino Oy.

Nummenmaa, L. (2004). Käyttäytymistieteiden tilastolliset menetelmät.
Mittaaminen ja aineisto, 34-37.
Tilastolliset testit ja hypoteesien testaaminen, 142, 143.
Helsinki: Kustannusosakeyhtiö Tammi.

Nunnally, J. & Bernstein, I. (1994). Psychometric theory. *The assessment of reliability,* 248-265. (3rd edition). New York: McGraw-Hill.

Sackman, H., Erikson, W. & Grant, E. (1968). Exploratory experimental studies comparing online and offline programming performance. *Communications of the ACM (CACM Homepage archive, 11* (1), 3-11.

Schmidt, F., Gast-Rosenberg, I. & Hunter, J. (1980). Validity Generalization Results for Computer Programmers. *Journal of Applied Psychology*, *65* (6), 643–661.

Schmidt, F. & Hunter, J. (2004). General Mental Ability in the World of Work: Occupational Attainment and Job Performance. *Journal of Personality and Social Psychology*, *86* (1), 162-173.

Shrout, P. & Fleiss, J. (1979). Intraclass Correlations: *Uses in Assessing Rater Reliability, 86* (2), 420-428.

Stanovich, K. (2004). How to think straight about psychology. *The role of change in psychology,* 164-173. *The Rodney Dangerfield of The Sciences: Our Own Worst Enemies,* 188-189. Boston: Allyn and Bacon, Pearson Education, Inc.

Vergauwe, E. & Cowan, N. (2014). Comparative A common short-term memory retrieval rate may describe many cognitive procedures. *Frontiers in Human Neuroscience, 8,* 1-7.

Viswesvaran, C., Ones, D. & Schmidt, F. (1996). Comparative Analysis of the Reliability of Job Performance Ratings. *Journal of Applied Psychology, 81,* (5), 557-574.

Wagner, R. & Sternberg, R. (1985). Practical Intelligence in Real-World Pursuits: The Role of Tacit Knowledge. *Journal of Personality and Social Psychology, 49,* (2), 436-458.

Witt, L. & Burke, L. (2002). Selecting High-Performing Information Technology Professionals. *Journal of End User Computing, 14* (4), 37-50.

ATTACHMENT 1: MINDFIT ANALYSIS 1/2

MindFit Consulting Oy

- Confidential -

Personal Analysis: John, born 1968

Analyzed by Kari Härkönen (12/2012)

Shortly

- education: Master's Degree in GIS and International Development (2001)
- work history: ..., Destia (2003-2008), UtilityMaps (2008-2009), Findzit (2010-2012)
- core knowledge: GIS, Requirements & functionalities of GIS
- management/leadership: experience in project management and team leading (5 – 15 members)
- communication/presentation skills: clear and business oriented style
- orientation/motivation: Development and architectures, intelligent and open team spirit
- special: *ESRI special achievement award in GIS services (2004)*
- profile: Development oriented architect (not on a code level) with business understanding
- salary: 5000-5500 €/kk or 80 €+/- as a consultant

Person Description

John has studied Business Administration and Geographical Information Systems (GIS) in Finland and in USA (Master's Degree from Clark University, USA, 2001). He has worked within GIS since 1989, e.g. many years in Egypt, few years in USA (Worcester City and Geophysics company, 2000-02), as well as many years in Finland having Destia's digital road assignments as his main projects. Much of his role has been to understand and specify the functionalities of the GIS, but he's also took coordination, management and team lead responsibilities as well as been involved in business operations. To be noted that their project won the ESRI special achievement award in GIS services. John's speciality is idea generation, requirement handling and definitions regarding GIS services. He knows the architectures and database principals (e.g. Oracle) in an overall level but is not a code level worker himself. Instead, he has a strong understanding of the GIS functionalities and he understands the business well, also from the cash flow / business model perspective. John knows the production phases well and can, for example, describe very clearly and determinately the business strategies from idea to implementation and markets (noted in a presentation test). John is, however, the most interested in creating new things and working with specialist team that appreciates intelligence and creativity. He could feel frustration in trivial tasks or in pure administrative assignments. As an enthusiastic student of the Geopolitics (~geographical affects to political decision making), John has combined the strategic and philosophic thinking to a very practical level ideating which makes him possible to participate in production in an implementation level.

Applicability

To work as a specialist or coordinator in an intelligent development team. Tasks that require definition, requirement handling and idea generation would suit him well. Can present products and processes from the functional side and from the business perspective. Can take part in implementation, but not on a code level. Could be valuable in supporting sales team, but not to work on pure sales or administrative roles. Could give valuable effort and knowledge to Start-up companies regarding market research and IPR processes.

ATTACHMENT 1: MINDFIT ANALYSIS 2/2

MindFit Consulting Oy

Competence List

Idx	Competence	From	Years	Months	Last Practiced
1	Project Management (3-9 in team)	Industry	3	0	0-6m
2	Project Mgmt (10-20 in team)	Industry	5	0	3-5y
3	Workshop Leading	Industry	10	0	0-6m
4	Team leading	Industry	6	0	0-6m
5	Windows XP (technical)	Industry	2	0	3-5y
6	Requirement Specifications	Industry	10	0	0-6m
7	Product Design specifications	Industry	10	0	0-6m
8	Implementation specifications	Industry	10	0	0-6m
9	Function specifications	Industry	10	0	0-6m
10	Business Development	Industry	4	0	0-6m
11	Oracle dBA experience	Industry	2	0	3-5y

ATTACHMENT 2: QUESTIONNAIRE FORM (NEW) 1/2

Candidate:	<First Name>, <year of birth>
Analysis:	<MindFit_Analysis_Name_mm_yyyy.doc>
Evaluator:	<First Name>, Last Name>

I assure that I have been in close cooperation with the employee for at least 4 months and that I have an objective understanding about his/her skills and competences in relation to requirements of the job tasks and expectations from the organization they are working.

<Name, title>, evaluator

Table 1 instructions:

Table 1: Evaluate the level of the competence (ability, skill, trait,...) of the candidate. Use the following 7 step (0-6) scale to evaluate the following 10 criteria. See Criteria Description and Level Description tables below to understand what each level value means for each criteria.

Table 2 instructions:

Table 2: Define how relevant such a skill or a trait is for the work tasks in IT field.

Table 1 **Table 2**

	Criteria	The level of criteria (ability, skill or trait)								Relevance of criteria at the work assignment		
	(Ability, skill, trait, appl.)	0	1	2	3	4	5	6		Very important	Good to know	Not relevant
1	Core skill											
2	Technical skills											
3	Communication skills											
4	Presentation skills											
5	Management skills											
6	Personal qualities											
7	Orientation											
8	Career expectations											
9	Risks (0 = high; 6 = no risk)											
10	Applicability											

ATTACHMENT 2: QUESTIONNAIRE FORM (NEW) 2/2

Criteria Descriptions:		Level Descriptions per Criteria:
Core skill = The very best skill over other knowledge and competence that the candidate has. For example, for the developer it could be a specific programming language. For some person the Project Management could be the very best skill to be evaluated here.		**Core skill, Technical skills:** 0 = No competence at all; **1** = Has weak understanding of the issue. Is not able to work independently with the task; **2** = Has basic knowledge of the issue. Can help in performing the task if getting strong guidance; **3** = Has an overall level understanding of the issue. Can perform the task rather indepently if getting support when needed; **4** = Good knowledge of the issue. Can perform the task independently and search the needed knowledge from the right sources; **5** = Excellent knowledge of the issue. Can perform the task independently and tutor others; **6** = Expert of the issue. Can improve the method itself, create course material and teach others.
Technical skills = The most relevant technical skills that the candidate has regarding the work in IT field (e.g. Automation testing, C++, ...).		
Communication skills = How well the candidate communicates with other people such as team members, colleagues, supervisors, customers, business people and other interest groups.		**Communication skills: 0** = Impossible to communicate with the person; **1** = Weak skills on communication; **2** = Basic skills on sharing thoughts and understanding others; **3** = Can take minor responsibilities regarding company's information sharing; **4** = Good skills on sharing, understanding and delivering information within interest groups; **5** = Excellent communication and information handling skills, can e.g. be responsible of company press releases/interviews; **6** = Expert skills on communication with all the needed interest groups (technical, business, public,...), plus can teach communication techniques to others.
Presentation skills = How well the candidate can share the information he/she has. This is to be evaluated considering sharing info within familiar people (such as own team members) as well as in a larger scale (such as keeping public / formal presentation).		**Presentation skills: 0** = No skills at all; **1** = Weak skills on presentation; **2** = Basic skills on sharing information to small team; **3** = Can make presentation on his/her own speciality area to small group; **4** = Good skills on sharing information to different target groups (e.g. technical and business); **5** = Excellent skills on drawing presentations in a very understandable way regarding the target audience needs; **6** = Expert on making presentations to any audience, can use different presentation techniques and teach those to others.
Management skills = How well the candidate can coordinate operative matters (e.g. from the technical team leading point of view) and manage larger groups (e.g. as a project manager). Consider also from the human leadership point of view.		**Management skills: 0** = No experience nor skills; **1** = No experience, some theoritical understanding; **2** = Basic skills, can e.g. help team leader in coordination; **3** = Can independently lead small team in his/her own competence area; **4** = Good skills on leading teams and small projects, knows the most common methods such as Agile/Scrum, Waterfall, has also human leadership qualities; **5** = Excellent skills on both, operative and human leadership roles, can motivate others; **6** = Expert leader, can coach and motivate others, helps team and organization to achieve challenging goals, and is able to teach
Personal qualities = How does the other personal qualities (e.g. interests, flexibility, attitude towards learning new things) support the adaptability of the practises within IT organization.		**Personal qualities: 0** = Not suitable at all; **1** = Has very few qualities that could fit to working with organization in the IT field; **2** = Has qualities that can help him/her to to learn practises in the IT field; **3** = Has potential to work to rather independently in the IT field; **4** = Good qualities regarding work assignments and possible changes in the procedures of the IT field; **5** = Excellent qualities regarding work assignments in the organizations of the IT field, can help organization to adapt changes in procedures; **6** = Expert qualities regarding any work assignment in the IT field, can teach organization manners and processes to others and prepare them to changes.
Orientation = How well the candidate is oriented to work in IT field in general and to develope his/her skills on the field.		**Orientation: 0** = Not oriented at all; 1 = Can do some IT tasks in a short term, has very little orientation or interest to work in the IT field in a long term; **2** = Is oriented to do short term projects in the IT field, no long term commitment (e.g. not over 6 months); **3** = Is oriented to work on IT field, some risks are related to long term commitment (over 1 year); 4 = Good orientation and strong interest to work in the IT field, no major commitment risks; **5** = Strong orientation to work in the IT field, has no major interests to move on other fields in a long term; **6** = Is a pure IT person, cannot see any other technology areas nor industry fields as a work place.
Career expectations = How well the candidate's career expectations fit to tasks and responsibilities required in IT field.		**Career expectations: 0** = has not planned to develop him-/herself regarding IT field; **1** = has very few thoughts about the career development in a long term; **2** = has figured out some career options for the future; **3** = knows own abilities and have rather realistic view about how to develop those to right direction; **4** = Good and organized plan for the future regarding the career; **5** = Excellent and organized career plan, flexible to develop it also regarding the organization needs; **6** = Expert in career planning, realizes own abilities and can see realistic possibilities to others, as well as teach others in their career planning.
Risks = Possible risks related to recruitment of the candidate or his/her ability to commit to projects (such as candidate's motivation in a long term). Level 0 means extremely high risk and level 6 refers to situation where no risks at all can be seen.		**Risks: 0** = Hiring the candidate is a huge risk! It is not recommended to be in any cooperation with the candidate; **1** = Cooperation with the candidate could be high risk for the organization, including personality and motivation factors; **2** = Hiring the candidate includes risks, such as motivation to commit in a long term; **3** = Only minor risks regarding short term project commitment (3-6 months), risk factor grows in a long term; **4** = No markable risks can be seen now or in a long term, if the work description and agreement terms follow the planned line; **5** = Excellent commitment, no mentionable risks regarding hiring nor any other cooperation; **6** = 100% trustable person in any circumstances, can be trusted when he/she presents company in public (e.g. media) or in a private discussions. Would never leave the organization in truble and helps the company to achieve the risk free status in their business.
Applicability = How well the candidate suits for the tasks and responsibilities required in an IT field in general. Note that even the expert candidate does not need to know everything about the field, but certain interdisciplinary can be helpful in learning new things.		**Applicability**: **0** = Not suitable at all; **1** = Has very few qualities that could fit to working with organization in the IT field; **2** = Has qualities that can help him/her to to learn technologies and processes in the IT field; **3** = Has potential to work to rather independently in the IT field; **4** = Good qualities regarding work assignments, changing requirements as well as the ability to learn new things; **5** = Excellent qualities to handle different work assignments and changing environment. Can adapt new things fast and help organization to develop their technical/business strategies; **6** = Expert qualities regarding any work assignment in the IT field, follows the cutting edge technical/business devlopment systems and methods and can adapt those to organization in order to maximize its competitiveness in a market.

Notes:

If some criteria was not mentioned in an analysis at all, the criteria will not decrease the reliability (will be left out from the statistical analysis no matter if it is relevant criteria or not).

ATTACHMENT 3: QUESTIONNAIRE FORM (OLD) 1/2

Candidate:	<First Name>, <year of birth>
Analysis:	<MindFit_Analysis_Name_mm_yyyy.doc>
Evaluator:	<First Name>, Last Name>

I assure that I have been in close cooperation with the employee for at least 4 months and that I have an objective understanding about his/her skills and competences in relation to requirements of the job tasks and expectations from the organization they are working.

<Name, title>, evaluator

Table 1: Evaluate the level of the comptence (ability, skill) of the candidate. Use the following 7 step scale in the first 5 criteria (1 to 5). Table 2: Define how relevant such a skill is for the work tasks in IT field:
0 = No competence at all.
1 = Has weak understanding of the issue. Is not able to work with the task.
2 = Has basic knowledge of the issue. Can help in performing the task if getting strong guidance.
3 = Has an overall level understanding of the issue. Can perform the task rather indepently if getting support when needed .
4 = Good knowledge of the issue. Can perform the task independently and search the needed knowledge from the right sources. Is capable of technical / operative coordination or leading small team.
5 = Excellent knowledge of the issue. Can perform the task independently and tutor others. As a manager he/she has both operative as well as human leadership skills.
6 = Expert of the issue. Can improve the method itself, create course material and teach others. Could work as a manager and director having extensive people skills and ability to motivate others.

Table 1: Evaluate the level of suitability (trait, applicability) of the candidate. Use the following 7 step scale in the last 5 criteria (6 to 10). Table 2: Define how relevant such a trait is for the work tasks in IT field:
0 = Not suitable at all. Hiring the candidate is a huge risk!
1 = Has very few qualities that could fit to working with organization in the IT field. Hiring the candidate could be high risk for the organization.
2 = Has some qualities that can help him/her to work at the organization in the IT field. Hiring the candidate includes risks.
3 = Has potential to work to rather independently in the organizations of the IT field. Hiring the candidate includes some risks.
4 = Good qualities regarding work assignments in the organizations of the IT field. Hiring the candidate includes very few risks.
5 = Excellent qualities regarding work assignments in the organizations of the IT field. Hiring the candidate is not a markable risk.
6 = Expert qualities regarding work assignments in the organizations of the IT field. Hiring the candidate is not a risk.

Table 1

	Criteria	The level of criteria (ability, skill or trait)						
	(Ability, skill, trait, appl.)	0	1	2	3	4	5	6
1	Core skill							
2	Technical skills							
3	Communication skills							
4	Presentation skills							
5	Management skills							
6	Personal qualities							
7	Orientation							
8	Career expectations							
9	Risks (0 = high; 6 = no risk)							
10	Applicability							

Table 2

Relevance of criteria at the work assignment		
Very important	Good to know	Not relevant

ATTACHMENT 3: QUESTIONNAIRE FORM (OLD) 2/2

Notes:

Explanations for the criteria to be evaluated:
Core skill = The very best skill over other knowledge and competence that the candidate has. For example, for the developer it could be a specific programming language. For some person the Project Management could be the very best skill to be evaluated here.
Technical skills = The most relevant technical skills that the candidate has regarding the work in IT field (e.g. Automation testing, C++, ...).
Communication skills = How well the candidate communicates with other people such as team members, colleagues, supervisors, customers, business people and other interest groups.
Presentation skills = How well the candidate can share the information he/she has. This is to be evaluated considering sharing info within familiar people (such as own team members) as well as in a larger scale (such as keeping public / formal presentation).
Management skills = How well the candidate can coordinate operative matters (e.g. from the technical team leading point of view) and manage larger groups (e.g. as a project manager). Consider also from the human leadership point of view.
Personal qualities = How does the other personal qualities (e.g. interests, flexibility, attitude towards learning new things) support the adaptability of the practises within IT organization.
Orientation = How well the candidate is oriented to work in IT field in general and to develope his/her skills on the field.
Career expectations = How well the candidate's career expectations fit to tasks and responsibilities required in IT field.
Risks = Possible risks related to recruitment of the candidate (such as candidate's motivation in a long term). Level 0 means extremely high risk and level 6 refers to situation where no risks at all can be seen.
Applicability = How well the candidate suits for the tasks and responsibilities required in an IT field in general. Note that even the expert candidate does not need to know everything about the field, but certain interdisciplinary can be helpful in learning new things.

Notes:

If some criteria was not mentioned in an analysis at all, the criteria will not decrease the reliability (will be left out from the statistical analysis no matter if it is relevant criteria or not).

ATTACHMENT 4: FORCED CHOICE METHOD

MindFit Consulting Oy

Forced Choice method, profile and instructions

Ask fast tempo questions like *'Which one you would be if you would have to choose?'* Also make crossing questions in order to trap possible faking effect. Put the result in Shortly and explain details in Person Description. Example of questions:

Roles:

- tester or coder
- coder or team leader
- architect or developer
- coder or architect
- tester or architect
- developer or project manager
- project manager or test engineer
- project manager or account manager
- team leader or sales person
- line manager or project manager
- technical project manager – or business oriented project manager
- sales manager or project manager

Tasks:

- definition or implementation
- requirement specification or test reporting
- coding or definition
- coding or testing
- development or team leading
- etc.

Choose the proper ones (not all necessarily, use different combinations and roles if needed based on job description and candidate, try to catch fake responses. After questions you can discuss about roles in detail.

Result examples:

- *technical oriented Line Manager with strong sales interest*
- *business oriented PM with technical communication skills*