Competencies of the Sexiest Job of the 21st Century: Developing and Testing a Measurement Instrument for Data Scientists

Human Resource Studies, Tilburg University, the Netherlands

Master Thesis on HR analytics, 2017

Author: Rosanne Bollen (U1250237)

Supervisor: Marinus Verhagen

Second reader: Paul van der Laken

Abstract

The aim of this research was to improve and validate a competency model for data scientists which can be used to predict their performance. The performance criterion was defined in task performance, organizational citizenship behavior (OCB), and counterproductive work behavior (CWB). In the first study 21 data scientists completed a self-report questionnaire about their competencies and performance, and they provided feedback about the competency model. Furthermore, supervisor performance ratings, official performance rating and general cognitive ability scores were analyzed together with the data from the competency questionnaire. No convergent validity was found between the three performance measurements. Also, no relationship was found between general cognitive ability and the performance self-reports. Study 2 attempted to validate the revised measurement instrument from study 1 through questionnaire data from 76 data scientists. Ten reliable competency domains were identified and compared with the original competency framework. Also, reliable short scales were composed to shorten the questionnaire. As a result, a new competency framework was proposed consisting of fewer subdomains with less items. Indications for the predictive validity of two competencies ‘planning & interaction’ and ‘business knowledge’ on OCB were found. These competencies were able to distinguish top from average performers. Moreover, the results indicate that the competencies are mostly related to OCB, less to task performance, and not at all to CWB. This study contributed to the validation of a competency model for a complex and rare profile, which is relevant for both practice and academics.

Keywords: Data scientists, competencies, competency model, individual performance, task performance, organizational citizenship behavior, counterproductive work behavior, general cognitive ability.

ACKNOWLEDGEMENT

This thesis would not have been possible without the support of my supervisor Marinus Verhagen. He provided me with honest, constructive, and useful feedback and his guidance was indispensable to achieve this result. I learned many things from his advice and I valued and enjoyed the consultation sessions. He allowed me to have much autonomy, but at the same time he steered me in the right direction when necessary. I also express my gratitude to my second reader Paul van der Laken for his enthusiasm about my research topic and for participating in the study.

I would like to express my appreciation for MIcompany, the organization that provided me with the unique opportunity to do this research, to make their data available, and to provide me with relevant contacts for this research. The assignment in combination with the autonomy I was given for this project motivated me to deliver real impact. In particular, I want to thank Madelon Otto, who enthusiastically started this project with me and guided me through this period. Also, special thanks go to Tryntsje Hoving-Wesselius, who took over the role of supervisor. Her knowledge and advice definitely helped throughout the project. Another person who deserves special thanks is Jonathan Koo, who critically assessed the competency domains with me. Furthermore, I want to thank all other experts from MIcompany who were involved in this research project by completing the questionnaire, providing feedback, sharing thoughts, and providing me with organizational data. Moreover, I am grateful for all managers who promoted the questionnaire with enthusiasm amongst their data scientists. On top of that, I wish to thank all data scientists who filled in the questionnaire and made the validation of the competency framework possible. Without the cooperation and energy of all these people, this research project could not have been completed.

Last but not least, I thank my family and friends, who frequently asked about my project and provided me with feedback, ideas, and positive energy. Their support has been motivating throughout the whole period.

Contents

Introduction

In collaboration with a marketing intelligence organization (organization X), this study aimed at developing and validating a competency framework for data scientists. Organization X had developed the framework based on their experience and with these competencies they claimed to have the right ingredients to identify and develop great data scientists.

This study contributes to science by developing and testing a competency framework on its predictive ability of individual performance. It is based on theory and had the unique opportunity to include organizational data such as intelligence scores and official performance ratings in its research. Also, the roles and competencies of the data scientist are currently rather unexplored, even though these are increasingly requested by organizations. Furthermore, this study was conducted in collaboration with an organization where data scientists work. Insights from these experts were used to revise the competencies and to check interpretations.

This thesis is relevant for practice, since it encourages the development and application of competencies in analytics in two ways. First, good data scientists are crucial for the organization and are often referred to as ‘unicorns’ which are rare (Willems, 2015). In McKinsey Global Institute’s report about big data, it is emphasized that using big data is key for an organization’s competitive advantage, as it can create value in many ways (Manyika et al., 2011). On LinkedIn ‘statistical analysis and data mining’ is ranked the second most searched skill globally, and first in the Netherlands (Fisher, 2016). Moreover, data scientist is even said to be the sexiest job of the 21st century (Davenport & Patil, 2012). Second, since this profile is important for organization, the recruitment, training and retention of data scientists are important as well. This is the crucial responsibility of the Human Resources department

(Armstrong & Taylor, 2014). Therefore, it is essential to know what makes a data scientist a good performer, and how this can be measured and predicted. With a validated measurement of a competency framework for data scientists, HR can help organizations to build analytical capability and thereby improve organizational performance (Manyika et al., 2011). Typically, HR possesses a great variety and amount of employee data (Cascio & Boudreau, 2011). However, HR appears to be the business area that uses the least business analytics (Services, 2011). It appears to have hit the “wall”, where it lacks the predictive capability and sufficient understandings of the variables and relationships (Cascio & Boudreau, 2011). If HR wants to contribute to strategic objectives, it will need to develop credible arguments through HR analytics (Fitz-enz, 2010). In short, HR analytics can be defined as an HR practice based on information technology to create business impact and enhance data-driven decision-making (Marler & Boudreau, 2016)**.** There is general consensus that analytical skills among HR professionals are lacking and inhibiting the progression in this field (Angrave, Charlwood, Kirkpatrick, Lawrence, & Stuart, 2016; Edwards & Edwards, 2016; Marler & Boudreau, 2016; Nicolaus et al., 2016; Russom, 2011). Therefore, Ulrich and Dulebohn (2015) argue that HR professionals need (analytical) competencies that deliver value. Organizations are thus in a high need of data scientists as well as analytical skilled employees for HR.

Before analyzing the validity of a competency model, the performance criterion needs to be clarified (Sackett & Lievens, 2008). Individual performance can be operationalized in various ways (Mayo, 2001). This study incorporates a behavioral and multidimensional definition where performance is divided into task performance, organizational citizenship behavior (OCB), and counterproductive work behavior (CWB) (Sackett & Lievens, 2008). The most important predictor of performance is general cognitive ability (GCA) (Smith & Smith, 2005) and is included in this study.

The aim of this thesis is to develop and validate a competency framework for data scientists to predict their performance. The main research question and subquestions are:

“*To what extent can the competency framework of data scientists be improved and validated in order to predict individual performance*?”

1. How can the individual performance of data scientists be defined and measured?
2. Which constructs are relevant predictors of performance and how can they be measured?
3. Which competencies may be relevant for data scientists?
4. With which scales can these competencies be measured?
5. How can the measurement instrument be shortened?
6. What is the predictive validity of the competencies on the performance of data scientists?

Theoretical Framework

The theoretical framework starts with the performance criterion, which is relevant for question 1. This is followed by performance predictors such as competencies and general cognitive ability, where an answer to question 2 is given. Then, the profile of the data scientist is discussed which is relevant for question 3.

The Performance Criterion

Performance is a widely researched concept which has received much attention in the fields of economics (Richard, Devinney, Yip, & Johnson, 2009), human resource accounting (Flamholtz, 2012), and organizational psychology (Spector, 2012). There is a lack of consensus about the criterion domain of performance and how it should be measured (Mayo, 2001). Therefore, a broad overview of these fields is provided. Subsequently, an outline of the antecedents and causal relationships of performance will be presented.

Economical perspective. From an economical perspective, performance is often referred to as organizational outcomes, including financial performance (e.g. ROI), product market performance (e.g. sales), and shareholder return (e.g. shareholder return) (Richard et al., 2009). However, this study field has barely paid attention to performance on the individual level (Mayo, 2001). Thus, the economical perspective is not considered appropriate for the individual performance of data scientists.

Human resource accounting. Another perspective which more specifically takes into account the individual level is called Human Resource Accounting (HRA) and is based on the economic theory of human capital (Flamholtz, 2012). It ought to quantify the value of individuals as a way of including employees as assets on the balance sheet (Flamholtz, 2012), for example, by adding historical expenditures made for the employee, such as training costs (Latham & Whyte, 1994). HRA has received a lot of criticism of opponents who state that human resources are no assets, that not every part of a person's job performance can be expressed in monetary terms (Latham & Whyte, 1994), and that accounting measurements are problematic for intangible resources (Steen & Welch, 2011). Some researchers therefore argue that HR accounting has reached a 'dead end' (Mayo, 2001). More meaningful metrics should be applied to evaluate the performance of individuals and to include perspectives from other research disciplines (Steen & Welch, 2011).

Organizational psychology. Much research about individual performance has also been done in organizational psychology (Kanfer & Kantrowitz, 2002). From an organizational psychology perspective, performance is behavioral (Campbell, McCloy, Oppler, & Sager, 1993; Viswesvaran, 2001). This study elaborates on two perspectives within this field.

First, some researchers tried to quantify individual performance by examining the consequences of behavior (Cascio & Boudreau, 2011). For example, the costs that arise because of turnover and absenteeism can be calculated (Cascio & Boudreau, 2011). This can be done through utility analysis, which aims to maximize utility and steering decisions by calculating the total costs and benefits of certain actions (Cascio & Ramos, 1986). Similar to the HRA approach, these methods of quantifying behavior do not capture the entire value of individuals and therefore the issue in identifying a person's value remains unresolved.

Second, other researchers argue that job performance is a multidimensional concept and can be clustered into three performance domains, namely task performance, organizational citizenship behavior (OCB) or contextual performance, and counterproductive work behavior (CWB) (Koopmans et al., 2011; Sackett & Lievens, 2008; Viswesvaran & Ones, 2000). Task performance refers to the application of core technical and task knowledge and determined by behavior that is specific for the job (Koopmans, Bernaards, Hildebrandt, de Vet, & van der Beek, 2014). The terms contextual performance and OCB can be used interchangeably and refer to extra-role behavior that is not stated in one's job description and contributes to the effectiveness of the organization (Ones & Viswesvaran, 2011). Counterproductive work behavior can be defined as deviant behavior that violates organizational norms and can threaten the well-being or effectiveness of the organization or stakeholders, and is thus an undesired form of performance (Marcus, Taylor, Hastings, Sturm, & Weigelt, 2016). Examples of such deviant behavior are theft, sabotage, and absenteeism (Marcus et al., 2016). Some researchers argue for a fourth dimension which is adaptive performance. This construct, in short, refers to dealing with change, uncertainty and problems, and the ability to learn from and adapt to them (Pulakos, Arad, Donovan, & Plamondon, 2000). However, a meta-analysis found that adaptive performance is not a separate dimension, but rather a part OCB (Koopmans et al., 2011).

An advantage of measuring performance in terms of behavior is that individuals have control over their own behavior (Viswesvaran, 2001) and all other factors that have an impact on the final outcomes are neglected. This approach can therefore be seen as more accurate on the individual level. A disadvantage is that measures of behavior are generally subjective and include rater biases (Viswesvaran, 2001). This will be elaborated upon further below.

This section provides an answer to the first part of research question 1: "How can the individual performance of data scientists be defined?" This study follows the organizational psychology perspective about the multidimensional and behavioral performance criterion.

Measuring Performance

To measure the three dimensions of individual performance, an accurate and valid measurement instrument is needed. Researchers have set different criteria for the assessment of a measurement's quality. According to Viswesvaran (2001) the six most common criteria are discriminability across individuals, practicality, acceptability, reliability, comprehensiveness, and construct validity. There is little consensus about a best measurement for individual performance (Koopmans, Bernaards, Hildebrandt, Van Buuren, et al., 2014). Below, a short review about performance measurements is presented.

Broadly, performance measurements can be divided into objective measurements (organizational data) and subjective measurements (Viswesvaran, 2001). Objective measurements are generally used for the economical approach to performance, HRA, or utility analysis. These objective data are generally free of rater biases that frequently occur with subjects measures (Viswesvaran, 2001). Nevertheless, criterion contamination and criterion deficiency can occur (Brogden & Taylor, 1950). The former occurs when the records are determined by external influence which are beyond control of the employees. The latter occurs when crucial elements of performance are not included in the objective measures. Although they can be practical, these types of measurements are generally low in acceptability, construct validity and reliability.

Subjective measurements, on the other hand, include ratings of employee performance (Viswesvaran, 2001). Employees' performance can, for example, be rated by supervisors only. As of the 90's, however, the popularity of the 360 degree or multisource feedback has increased tremendously, where different persons evaluate the employees, such as the employees themselves, supervisors, subordinates, colleagues, and clients (Atwater, Waldman, & Brett, 2002). The more types of raters are combined, the more biases can be overcome. This can increase reliability and acceptability, but may be less practical. Also, several rater errors and biases can occur with subjective evaluation, such as centrality bias (the tendency to compress performance ratings) and leniency bias (the tendency to inflate performance ratings) (Bol, 2011). Also, relative rating scales, where the performance is compared to performance of colleagues, are thought to provide better criterion validity results than absolute rating scales (Goffin, Gellatly, 1996). They force the rater to discriminate which helps to overcome biases such as centrality. When these biases are taken into account and prevented as much as possible, it is possible to capture the multidimensional performance of employees. Another way of rating employees is through rating systems. The most widely known rating systems are graphical rating scales (GRS) and behaviorally anchored rating scales (BARS) (Viswesvaran, 2001).

When designed correctly, they, and in particular BARS (Debnath, Lee, & Tandon, 2015), can be valid and reliable measurements (Viswesvaran, 2001). Nonetheless, to be able to use these systems, a lot of time and training is required.

From previous paragraphs an answer can be given to the second part of research question 1: “How can the individual performance of data scientists be measured?” Rating systems such as BARS appear to be the best measurements for performance. When such a system is not used, it is best to include different (types of) raters. A variety of subjective measurements can be applied to measure individual performance and the researchers or organization should strike a balance between the criteria mentioned earlier.

Predicting Performance

Below, the literature about the antecedents and causal relationships of performance will be presented, which will provide an answer to research question 2. On the individual level, research about performance pays much attention to individual differences (Viswesvaran, 2001). Generally, the individual differences are split up into ability and non-ability predictors of job performance (Kanfer & Kantrowitz, 2002). Ability predictors often refers to general cognitive ability (GCA) (Hunter & Schmidt, 1996). Next to GCA, competencies are often defined as ability (Boyatzis, 2008) and will therefore be considered as an ability predictor of performance. The non-ability predictors often include personality and contextual factors (Kanfer & Kantrowitz, 2002; Paauwe & Boselie, 2005; Viswesvaran, 2001). Personality is generally found to best predict OCB (Sonnentag & Frese, 2002). Other examples of performance predictors include engagement (Rich, Lepine, & Crawford, 2010) and well-being (Peccei, van de Voorde, & Van Veldhoven, 2013). Although these variables can have influence on performance, they are out of scope for this study. First, competencies are elaborated upon and the link with performance will be explained. Thereafter, general cognitive ability is discussed.

Competencies

Central to this research are competencies. Competencies were introduced by psychologist McClelland (1973), who argued that competencies are better predictors of occupational success than intelligence tests. The use of competencies in organizations has increased in popularity, and it has largely replaced the job task analysis (Cardy & Selvarajan, 2006; Kurz & Bartram, 2002; Shippmann et al., 2000). An example of a competency model is the Great Eight (Bartram, 2005), which could be seen as a competency variant of the Big Five for personality (Bartram, 2005).

Despite its popularity, there is confusion about the definition of competencies (Kurz & Bartram, 2002; Shippmann et al., 2000). For example, competencies are defined as a “capability or ability” (Boyatzis, 2008, p. 6), others define it as “the knowledge, skills, and attributes that differentiate high performers from average performers” (Shippmann et al., 2000, p. 706), or “sets of behaviors in the delivery of desired results or outcomes” (Bartram, Robertson, & Callinan, 2002, p. 7). A good definition of competencies is important, since the construct must be specified before the construct measurement can be developed and validated. This study blends the last two definitions into: “The knowledge, skills, and behaviors that differentiate high performers from average performers in the delivery of desired results.” First, this study argues that competencies include knowledge, skills, as well as behavior. Second, competencies should be relevant to performance and therefore able to distinguish among average and excellent performers. Third, and in line with the second argument, competencies should be directed towards a desired result. Finally, competencies can be improved.

Competencies and performance. Using competencies to predict performance has the advantage that the link between the predictor and the criterion behavior is direct (Altink & Verhagen, 2002). Competencies are not the same as performance, but they are an enabler for performance (Kurz & Bartram, 2002) and can serve as a performance predictor (Markus, Thomas, & Allpress, 2005). Employees can have competencies while performing bad, but employees cannot perform well without having the necessary competencies. Furthermore, since work is becoming more complex, skill requirements are increasingly important (Markus et al., 2005). Spencer and Spencer (2008) argue that in complex jobs, competencies are more predictive of job performance than intelligence. This is primarily due to a restricted range in cognitive ability, since employees are required to have a high IQ (Spencer & Spencer, 2008). Finally, competencies can benefit HR since they provide guidance in practices such as recruitment and performance management (Markus et al., 2005).

There are, however, important validity issues with the competency approach (Markus et al., 2005). First, because of lack of agreement on the term ‘competency’, construct and content validity are hard to reach (Markus et al., 2005). Moreover, this lack of agreement has created confusion regarding the distinction between competencies and performance. When competencies are measured in terms of behavior, and when performance is defined by behavior as well (Campbell et al., 1993), there may be no clear distinction between the two concepts (Markus et al., 2005). This can lead to circularity in reasoning. Second, measuring competencies is hard, because it usually includes subjective ratings that suffer from various biases. Moreover, it is often assumed that using competencies is related to improved performance, although there is not much evidence (Markus et al., 2005). Finally, using a competency model which has not been validated is risky, because a competency model can promote certain behaviors, and inappropriate or irrelevant behaviors may be included if the competencies are not carefully analyzed (Shippmann et al., 2000). These drawbacks need to be taken into account when researching and applying competencies.

Developing and measuring competencies. For the effective performance in a job, a variety of competencies is needed, resulting in a competency model (Cardy & Selvarajan, 2006). A competency model can be created for a specific job which can be helpful when this job is critical to an organization's success and requires distinct competencies (Cardy & Selvarajan, 2006; Mansfield, 1996; Shippmann et al., 2000). Also, a model can be created for a larger group of jobs (Mansfield, 1996), for example the Great Eight (Bartram, 2005). Furthermore, competencies must be measurable (Shavelson, 2010). For example, employees could evaluate their own competencies through self-reports. Also, trained raters or supervisors can observe the employees and assess their performance and competencies (Shavelson, 2010). As with performance, biases and rater errors can occur when measuring competencies.

General Cognitive Ability (GCA)

There is a general consensus in research that general cognitive ability (GCA) is the most valid predictor of job performance (Kanfer & Kantrowitz, 2002; Schmidt & Hunter, 1998; Smith & Smith, 2005). GCA can be defined as "the relative speed and accuracy with which the brain processes complex information" (Smith & Smith, 2005, p. 23) and the ability to learn and acquire knowledge (Hunter, 1986; Schmidt & Hunter, 1998). Spearman (1904) discovered that intelligence is a general ability.

GCA and performance. GCA tests are found to have an average predictive validity (*r*) of .51 for individual job performance (Schmidt & Hunter, 1998). There are many theories and explanations for the high correlation between GCA and performance, such as job complexity (Kanfer & Kantrowitz, 2002), the three-stratum hierarchy (Carroll, 1993), performance model of Campbell et al. (1993), and the Gf-Gc (fluid and crystallized) model (Cattell, 1963).

This research elaborates on the classical theory of performance (Hunter & Schmidt, 1996), as this theory provides the best explanation between general cognitive ability and individual performance. It argues that both a direct effect, and an important indirect effect via learning exist. Especially in new situations that arise on the job, the direct relationship between GCA and performance is strong, because these new situations require adaption (Hunter & Schmidt, 1996). People must quickly link the information from the current situation with the knowledge they already have, evaluate the information and consequently make decisions for a response. All these activities require cognitive ability (Hunter, 1986). The higher the GCA, the higher the speed and accuracy with which people can process complex information (Smith &

Smith, 2005). Because in almost all jobs time and amount of information matter, the argumentation about the relationship between cognitive ability and performance can be generalized (Schmidt, 2002).

The classical theory also argues for an indirect effect via learning and the acquisition of job knowledge (Hunter & Schmidt, 1996; Thorndike, Bregman, Cobb, & Woodyard, 1926). People with high GCA are better able to quickly learn new job knowledge and tasks, and therefore perform high (Smith & Smith, 2005). These people thus process more information and they do it faster. Consequently, it is likely that a higher GCA enables higher performance through learning and the acquisition of job knowledge.

Measuring GCA. For measuring general cognitive ability there are many tests available. Most tests include measurements of general, verbal, and spatial intelligence (Smith & Smith, 2005). GCA is found to be relatively stable over time (Smith & Smith, 2005). For example, in a retest of the Moray House Test after 66 years a correlation of .73 was found (Deary, Whalley, Lemmon, Crawford, & Starr, 2000).

To conclude, this section answers research question 2: "Which constructs are relevant predictors of performance?" General cognitive ability is the best valid predictor of performance. This research is primarily concerned with competencies and takes GCA into account. There are also non-ability predictors, but these are out of scope for this study.

The Data Scientist

In previous paragraphs, literature about performance, competencies, and GCA was discussed. The aim of this research is to develop and test a competency measurement instrument of data scientists. Therefore, this study now turns to the data scientists. Below, some perspectives and empirical studies about data scientists are summarized and discussed.

A clear definition or job description of a 'data scientist' is hard to find. According to some, data scientist is only a buzzword for the term business analytics (Press, 2013). Others regard data scientists as a self-contained function. A clear overview of attempts to describe and conceptualize the profiles of data scientists in Venn diagrams was outlined by Taylor (2016). A few of them are briefly discussed.

To start with, Conway (2013) designed a Venn diagram containing three overlapping areas, which are hacking, math and statistics knowledge, and substantive expertise (referring to knowledge and theories concerning the topic of interest), where the data scientist is in the center. Harris (2013) altered the diagram to a 'data products Venn diagram' with three skillsets; domain knowledge, software engineering, and finally statistics, predictive analytics and visualization. The main difference lies in the addition of predictive analysis and visualization. In response, Kolassa (2016) addressed the overlooked importance of communication and argued for four areas, which are business, programming, statistics, and communication. An overview of these categorizations can be found in Table 1. These diagrams are similar in the sense that they depict a blend of different disciplines. Furthermore, all approaches include the fields of statistics, programming, and knowledge about the business or the field of interest. The main differences concern the predictive analyses, visualization and communication skills, and the number of areas. These diagrams are not tested empirically. Hence, the next paragraphs discusses empirical findings on the competencies of data scientists.

Empirical Findings

Allegedly the most cited researcher about data scientists is Davenport, since many other authors refer to his articles to describe data scientists (e.g. Power, 2014; van der Aalst, 2014). In 2001 he and colleagues presented a study in 20 companies, where five key skills were identified that would help building strong analytical talent (Davenport, Harris, de Long, & Jacobson). These were technology skills, statistical modeling and analytical skills, knowledge of the data, knowledge of the business, and communication and partnering skills. Also, a distinction was made between four analytical key roles: the database administrator, the business analyst or data modeler, the decision maker, and the outcome manager. The skills are relevant for all analytical roles, but the degree of relevance varies (Davenport et al., 2001). Later, Davenport and Patil (2012, p. 73) introduced the data scientist as the ‘sexiest job of the 21st century’. They defined a data scientist as “a hybrid of a data hacker, analyst, communicator, and trusted adviser” (Davenport & Patil, 2012). The data scientist can thus be seen as the umbrella role of the four roles as described by Davenport et al. (2001). Davenport (2012) also labeled the domains as technical, business, analytical, and relationship, which are almost identical to categories of the Venn diagram by Kolassa (2016).

Datacamp, an online data science education platform, described data scientists as rare unicorns who have skills distributed computing, predictive modeling, story-telling and visualizing, and finally in math, stats, and machine learning (Willems, 2015). Although the author attempted to distinguish the data scientist from other roles like the data analyst, she admits that no general definition exists of the data science roles. The skills show large resemblance across the different roles. The main differences with previous categorizations is that business or domain knowledge was excluded.

In another study which analyzed the competencies of business intelligence (BI) and big data (BD) professionals (Debortoli, Müller, & vom Brocke, 2014), the authors mentioned that data scientists are similar jobs to BD professionals. For BD jobs, two broad areas were found, namely business and IT competencies, where business was subdivided into domain and management, and IT into concepts and methods, and programming. The authors did not specify the similarities and differences between BD and data scientist jobs.

Next to that, a research studied the competencies of Business Intelligence and Analytics (BI&A) (Chiang, Goes, & Stohr, 2012). BI&A jobs and are defined as “an interdisciplinary area that integrates data management, database systems, data warehousing, data mining, natural language processing (..), network analysis/social networking, optimization, and statistical analysis” (Chiang et al., 2012, p. 3) which resemble data scientist roles. The competencies were subdivided in analytical, IT knowledge, and business knowledge and communication. Again, this distinction resembles the categorizations by Davenport (2012) and Kolassa (2016).

Previous described literature does not give a full answer to research question 3: “Which competencies are relevant for data scientists?” This shows the absence of proper empirical support for the competencies of data scientists and the ambiguity of the term data scientist, because this profile appears to be similar to other data science industry roles (Willems, 2015), the BI&A and BD jobs (Chiang et al., 2012; Debortoli et al., 2014), and they are not clearly distinguished. All researchers seem to agree that data scientists are multidisciplinary in at least three domains that refer to programming or technology, statistics, and business or domain knowledge. Next to that, communication and visualization are often highlighted as important. There remains disagreement about the competencies. The findings are summarized in Table 1.

Table 1 The development of the competencies domains of data scientists from Venn diagrams and empirical research

Authors	Technology	Analytics	Business	Communication
Conway (2013)[1]	Hacking skills	Math & statistics knowledge	Substantive expertise	
Harris (2013)[1]	Software engineering	Statistics, predictive analytics, visualization	Domain knowledge	
Kolassa (2016)[1]	Programming	Statistics	Business	Communication
Davenport (2012)[2]	Technical	Analytical	Business	Relationship
Willems (2016)[2]	Distributed computing	Predictive modeling	Math, stats, machine learning	Story-telling and visualizing
Chiang et al. (2012)[3]	IT knowledge	Analytical knowledge	Business knowledge	Communication
Debortoli et al (2014)[3]	Programming	Concepts and methods	Business domain	Management

Note. [1]Venn diagrams about data scientists. [2]Empirical evidence about data scientists. [3]Empirical evidence about jobs that are similar to those of data scientists.

Model of the Predictor-Criterion Relationship

For developing and validating the competency framework, the model in Figure 1 was used (Arthur Jr & Villado, 2008). First, the criterion needs to be specified and predictors are chosen based on theory (a). Individual performance are defined through a multidimensional and behavioral perspective. This study therefore refers to performance as task performance (TP), organizational citizenship behavior (OCB), and counterproductive work behavior (CWB). Second, GCA has been proven as a valid predictor of performance and is therefore included in study 1. It is likely that data scientists generally score high on GCA, since they fulfil complex jobs. Because of expected low discriminability, the predictive validity of competencies for data scientists may be higher. Third, data scientists are thought to have multidisciplinary competencies, but the exact categorization remains unclear.

The competency framework needs to be improved and validated. Valid measurement instruments need to be selected for the predictors (b) and the criterion (c). Finally, the predictor measurements ought to predict the criterion construct (e). However, only the validity between the measurements (d) can be measured. The more valid the measurements (b,c), and the stronger the theoretical foundation (a), the higher the predictive validity of the predictor measurements. More details about the studies can be found in the method section.

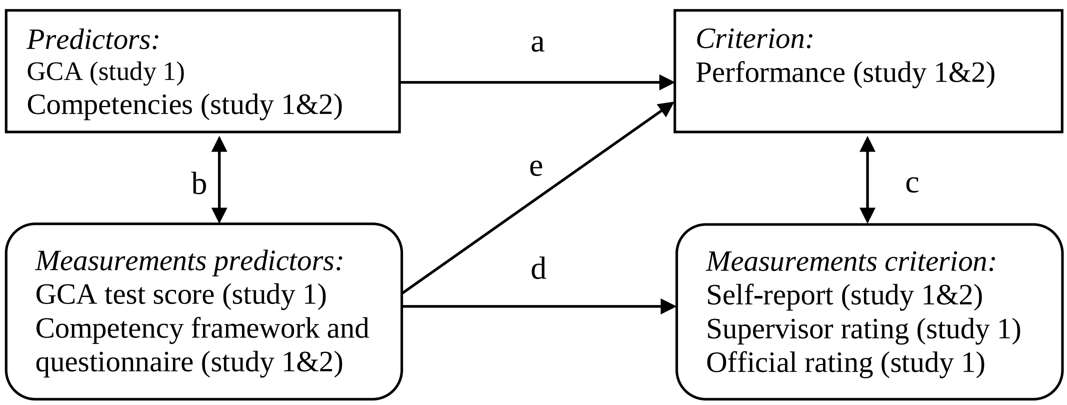

Figure 1Model of the predictor - criterion relationship for the improvements of the competency framework and questionnaire.

Method Study 1

The marketing intelligence organization wanted to stay anonymous and will be referred to as ‘organization X’. Organization X had developed a competency framework based on experience, which has not been validated. In order to develop and validate the competency measurement instrument, two studies were be conducted which will be described below. The population of interest for this thesis concerns both Dutch and international data scientists.

Research Design

First, study 1 was conducted to do a preliminary validation and to improve the competency framework and the questionnaire. Organizational data included GCA scores and official performance ratings from two moments in time. All data scientists were asked to complete a questionnaire about their competencies and performance. Additionally, supervisors were asked to rate the performance of their employees. With three measurements of performance, the convergent validity can be measured, which contributes to the answer for research question 1 about the measurement of performance. Analyzing GCA and performance gives clues about important predictors of performance in research question 2. Next to that, because organization X claimed that the competencies from the framework enhance performance, analysis was performed to check whether data scientists with high scores on competencies actually had high performance ratings. Results from this part provided some answers to research question 6 about the predictive validity of competencies on performance of data analysts. Next to this quantitative part, study 1 contained qualitative elements. The respondents provided feedback on the questionnaire to improve face and construct validity. This feedback helped to partially answer research question 3 and 4 about the relevant competencies and its measurement instrument.

Sample

This study used a convenience sample with data scientists and their supervisors from organization X, primarily because of their expertise, accessibility, and available organizational data. Since it is a small organization with 38 data scientists, only a small sample size could be obtained and no demographic information was asked. 21 data scientists responded to the questionnaire of which 8 were working students, and thirteen were junior or senior data scientists. All respondents were Dutch. Additionally, one English native speaker was interviewed for the English translations. The supervisors rated the performance of eight data scientists. Six working students and eight data scientists agreed to merge their data. Only four supervisor ratings, five official ratings, and ten GCA scores could be merged with the self-reports. This sample did not provide a representative sample for the population, but generalization was not the purpose of this first study.

Instruments

Performance. The performance was evaluated through self-ratings, supervisor ratings, and official ratings. For the self-ratings, the validated Individual Work Performance Questionnaire (IWPQ) (Koopmans, Bernaards, Hildebrandt, Van Buuren, et al., 2014) was used which can be found in Appendix 2. It included the three dimensions task performance (TP), organizational citizenship behavior (OCB), and counterproductive work behavior (CWB). Reliability scores of the scales were .78, .85, and .79 respectively (Koopmans, 2015). Task performance consisted of five items, OCB of eight items, and CWB of five items. The items were rated on a 5-point rating scale from seldom to always for task performance and OCB, and from never to often for CWB. An example question for task performance was: “I was able to plan my work so that I finished it on time”, for OCB: “I took on challenging work tasks when they were available”, and for CWB: “I made problems at work bigger than they were.” All items for CWB were recoded as they were negatively stated. Higher scores on CWB therefore means higher performance. Worth noting is that the CWB items do not seem to cover the actual CWB concept. CWB refers to deviant behavior, such as theft, sabotage, and absenteeism (Marcus et al., 2016). However, this CWB scale is rather a mild form and includes items about complaining about work.

Supervisor performance ratings. To reduce issues from the self-report biases, supervisors were asked to rate data scientists’ performance. The same questions from the IWPQ (Koopmans, Bernaards, Hildebrandt, Van Buuren, et al., 2014) were used and changed slightly only to change the personal pronoun from first- to third-person.

Official performance ratings. Official ratings for the junior and senior data scientists were already available for December 2016 and June 2017. They had first been rated by their supervisor, and the ratings were then discussed within the management team to reach rater consistency. The ratings ranged from 1 (poor) to 5 (excellent). As discussed before, evidence was found that relative performance evaluations provide better criterion validity than absolute evaluations (Goffin, Gellatly, Paunonen, Jackson, & Meyer, 1996).

General cognitive ability. The GCA scores were measured through an online test during their assessment in the selection procedure. The test included numerical series, figure series, and verbal analogies. The maximum score was 300, where 135-209 indicated average academic, 210-273 beyond average academic, and >274 far beyond average academic level.

Competencies. The competency questionnaire, developed by organization X, was used. It concerned a self-rating scale and it consisted of four general competency domains which were subdivided into subdomains. An overview can be found in Appendix 2. The general domains were data and technology (DT), analytical methods and techniques (AMT), impact and advisory (IA), and business domain expertise (BDE). In total, the original competency questionnaire contained 108 items. The structure, scales and face validity were first checked by the researcher in deliberation with a senior data scientist. From previous research about data scientists’ competencies, relevant items were added, and ill-defined items were adjusted or removed. An external trainer reviewed the structure and items of the impact and advisory skills. The final amount of items was 112. Items were rated on a 7-point Likert scale ranging from “beginner” to “expert”. This response scale was chosen to decrease the anchor problem where respondents have different interpretations of the scale (Allen & van der Velden, 2005), and to stimulate discriminability (Preston & Colman, 2000). Each question was preceded by: “Indicate to what extent you possess these competencies.” An example question for the data and technology domain was: “Programming the right results”, for analytical methods and techniques: “Interpreting the results correctly”, for impact and advisory: “Presenting in a confident and convincing way” and for business domain expertise “Measuring and monitoring the effect of an action/initiative.”

Feedback. Because the competency framework was developed based only on experience, the face and construct validity were reviewed by the data scientists. The respondents reviewed the content and the formulation of items, competencies that were missing, the eight most and the two least important competencies that excellent performers need, and the structure of the domains and items.

Procedure

The questionnaires were electronically distributed via Jotform.com with an explanatory message attached. Confidentiality was ensured and permission to merge the organizational data and performance ratings was asked. The questionnaires were in Dutch, except for one English version for an English native speaker. This person was interviewed to reformulate the English items if necessary and to check for consistency in interpretations. Assessment scores and the official ratings were collected via the HR officer. The data was merged and anonymized by the HR officer.

Analysis

Because the small sample size would seriously constrain the quality of factor analysis, the variables were correlated and analyzed through independent sample t-tests to identify differences between higher- and lower-performing data scientists. Also, these t-tests can identify differences between the three performance measurements, which include self-ratings, supervisory ratings, and official ratings. All analyses were performed in SPSS. Also, the feedback was analyzed and used to adjust the competency questionnaire for study 2.

Results Study 1

The results from study 1 consist of three parts. First, analyses are presented of GCA and the three performance measurements. Second, relationships between the competencies and performance is analyzed. Third, feedback on the questionnaire will be discussed.

RQ 1. Performance Measurements

When comparing the norm scores for white collar workers with the self-reports and supervisor ratings, task performance and CWB were relatively high and OCB was rated very high (Koopmans, 2015). Correlations were computed between the self-rated performance, supervisor ratings, official ratings, and the scores from the GCA test. No correlations between the different performance measurements of self-ratings, supervisor ratings, and official ratings were found that could provide evidence for convergent validity of performance measurement instruments. Significant correlations were only found between the supervisor ratings for task performance and OCB (r =.825, p =.012), and a negative correlation between the self-reports OCB and the supervisor ratings for CWB (r =-.978, p =.022). These correlations are high and need to be interpreted with caution, because of the small sample size (N=8 and N=3 respectively). Correlations can be found in Table 2.

Table 2 Correlations and descriptives: GCA, performance rating, self-report, supervisor report

		N	Mean	SD	1	2	3	4	5	6	7	8
1	GCA[1]	10	259.9	26.9								
2	Ratingdec	5	3.60	0.42	-.242							
3	Ratingjun	8	3.69	0.53	.184	.802						
4	TP[2]	21	3.32	0.54	.293	.071	.113					
5	OCB[3]	21	3.55	0.96	-.406	-.353	-.117	.406				
6	CWB[4]	21	4.10	0.46	.029	-.802	-.360	.236	-.138			
7	TPsup[5]	8	3.28	.72	-.877	[a]	-.132	-.118	.887	.466		
8	OCBsup	8	3.60	.74	-.694	[a]	.333	-.298	.586	-.236	.825*	
9	CWBsup	8	4.10	.62	.612	[a]	.554	-.297	-.978*	-.705	-.633	-.463

Note. *p<.05. [1]General cognitive ability. [2]Task performance. [3]Organizational citizenship behavior. [4]Counterproductive work behavior. [5]sup = supervisor rating. [a]No correlations due to missing values.

Since no convergent validity was found, paired sample t-tests were conducted to analyze whether there were significant differences between the performance scores from the self-reports with the supervisor ratings. No significant differences were found. In other words, the supervisors and data scientists did not disagree significantly about the performance. The paired sample t-tests can be found in Table 3.

Table 3 Paired sample t-tests between performance self-reports and supervisor rating

	Mean(dif)	SD	t	df
TP - TPsup	.050	.755	.132	3
OCB - OCBsup	.281	.544	1.035	3
CWB - CWBsup	.050	1.012	.099	3

Note. $^*p < .05$. N=4.

RQ 2. GCA as a Predictor of Performance

No significant correlations were found between GCA and the performance indicators. Furthermore, independent sample t-tests were performed to check whether data scientists with a GCA >270 would score higher on performance than data scientists with GCA <270. The results showed that there were no significant differences. Although expected, this indicated that data scientists with a higher intelligence did not score higher on any of the performance dimensions. Predictive validity of general cognitive ability could not be determined. Correlations can be found in table 2 and independent sample t-tests in Table 4.

Table 4 Results of independent sample t-tests for self-rating performance dimensions and GCA <270 and >=270.

	GCA <270[4]		GCA >=270			
	M[1]	SD[2]	M	SD	t	df
TP	3.28	.36	3.56	.17	1.565	5.6
OCB	3.80	.62	3.55	.77	-.565	8
CWB	4.24	.26	4.28	.46	.169	8

Note. $^*p < .05$. N=5. [1]Mean. [2]Standard deviation. [3]Number of respondents. [4]GCA 210-270 = beyond average academic, GCA >=274 = far above average academic.

RQ 3&4. Feedback on Competencies and Items

Eleven respondents filled in eight most and two least relevant items. To identify the most relevant items, the minimum amount of respondents per item was set at four, and for the least relevant items at two. With seven respondents, the most relevant item was “Translating the output into impact that is relevant for the business.” In other words, data scientists should be able to make a connection between analyses and the implications for the business. With three respondents, the less relevant item was “Defining and presenting personal vision on data & technology.” This most likely implies that either having a personal vision is found to be less relevant, or the item was not well understood. These results are presented in Table 5.

Table 5 Feedback on questionnaire. Most and least relevant items.

Item #	Most relevant items	# Respondents
33	Translating the output into impact that is relevant for the business	7
13	Translating the business question to the right analytical approach	5
21	Signaling opportunities proactively and quantifying potential	5
29	Asking questions and going in depth to understand the priorities and expectations	5
15	Interpreting results correctly	4
25	Translating insights from analyses to a relevant message	4
Item #	Least relevant items	# Respondents
23	Defining and presenting personal vision on data & technology	3
41	Premise shared goals above personal goals	2

The rest of the feedback was analyzed and adaptions were made when appropriate. For example, the first three option of the scale were changed to 1=no experience, 2=beginner, 3=intermediate, and the items of the data and technology competency were categorized more meaningfully. All feedback, considerations, and adaptions can be found in Appendix 3.

RQ 6. Predictive Validity of Competencies on Performance

Factor analysis was not possible due to the small sample size (N=21). The competencies were therefore computed through mean scores for general competency domains and correlated with task performance (TP), OCB, and CWB. Competencies from data & technology (DT) and analysis methods and techniques (AMT) were split into a general and specific variable, because they cannot simply be combined. All competencies will be discussed more in detail in study 2.

Positive significant correlations were found between task performance and general DT, specific DT, general AMT, and impact and advisory (IA). Scoring higher on these competencies was thus associated with higher task performance. OCB positively correlated with general DT, general AMT, impact & advisory, and business domain expertise (BDE). This indicates that higher scores on these competencies are positively related to the OCB of data scientists. CWB did only correlate with any variable. Descriptives and correlations can be found in Table 6.

Table 6 Descriptives and correlations of performance and competencies

		Mean	SD[1]	1	2	3	4	5	6	7	8
1	TP[2]	3.32	0.54								
2	OCB[3]	3.55	0.96	.406							
3	CWB[4]	4.10	0.46	.236	-.138						
4	DTgeneral[5]	4.52	1.35	.665**	.444*	.030					
5	DTspecific	2.12	0.48	.677**	.334	.365	.733**				
6	AMTgeneral[6]	4.43	1.16	.580**	.464*	.130	.412	.438*			
7	AMTspecific	2.79	0.89	.385	.094	.092	.147	.306	.715**		
8	IA[7]	3.97	1.07	.599**	.640**	-.058	.623**	.527*	.658**	.309	
9	BDE[8]	2.89	1.14	.389	.495*	-.045	.514*	.601**	.647**	.563**	.720**

Note. *$p<.05$, **$p<.01$. [1]Standard deviation. [2]Task performance. [3]Organizational citizenship behavior. [4]Counterproductive work behavior. [5]DT=Data & technology. [6]AMT=Analysis methods and techniques. [7]IA=Impact and advisory. [8]BDE=Business domain expertise.

Method Study 2

Research Design

In the second study, a quantitative cross-sectional questionnaire research was conducted in order to test the validity of the revised competency framework on a larger and more representative sample. To ensure a proper response rate, GCA and supervisor ratings were excluded. First, the scales of the competencies needed to be analyzed, since there was uncertainty about which items belong together in the same scale. This part will consequently provide insights in the competencies that arise from the analysis. Therefore, this will provide answers to research questions 3 and 4. Next to that, the established scales will be shortened to provide a more practical and less time consuming measurement instrument, hereby answering research question 5. Finally, the predictive validity of the competencies on the performance of data scientists will be analyzed, providing an answer to research question 6.

Sample

219 data scientists were approached through organization X. Although their official function titles varied, they were selected by organization X as ‘data scientist’. The questionnaire was also posted on LinkedIn. Two respondents were excluded because of an irrelevant function title. The final sample size was N = 76 of which 57.9% was male. Most respondents were between 26-30 years old (39.5%), followed by 31-35 (23.7%). The majority held a master’s degree (68.4%), followed by a bachelor’s degree (10.5%). The most mentioned function title was ‘data analyst’ (14.5%), but the largest category was ‘other analyst’ (39.5%) such as ‘pricing data analyst’. 11.8% labeled themselves as ‘data scientists’. 3.9% was categorized as ‘other’, for example ‘senior specialist’. The demographic characteristics can be found in Table 7.

Table 7 Demographic characteristics

Characteristic	Description	N	%
Gender	Male	44	57.9%
	Female	32	42.1%
Age	<26	8	10.5%
	26-30	30	39.5%
	31-35	18	23.7%
	36-40	9	11.8%
	41-45	4	5.3%
	46-50	7	9.2%
Education	High school/vocational	4	5.3%
	Associate degree	7	9.2%
	Bachelor's degree	8	10.5%
	Master's degree	52	68.4%
	PhD/other advanced degree	5	6.6%
Function	Data scientist	9	11.8%
	Data analyst	11	14.5%
	Other analyst	31	40.8%
	Managing function	10	13.2%
	Controller	5	6.6%
	BI	4	5.3%
	Other	3	3.9%
	Trainee	1	1.3%
	Missing	2	2.6%

Note. N=76.

Instruments

Competencies. The revised questionnaire was used in study 2. In Appendix 4 this questionnaire can be found. One open question regarding missing competencies was included. The rating scale was adapted and ranged from "no experience" to "expert". Since some organizations employed international people, the questionnaire was also distributed in English.

Performance. The same IWPQ scales were used for performance. PCA was performed for the performance items, since the factors were already known (Koopmans, Bernaards, Hildebrandt, de Vet, et al., 2014). Reliability scores (Cronbach's α) were .706 for task performance, .824 for OCB, and .840 for CWB.

Control variables. The control variables included age, gender, and education. These are common control variables and therefore included.

Procedure

Managers from client organizations were asked for contact details of data scientists and permission to contact them. Confidentiality was ensured and the data was not visible to the client organization nor organization X. A reminder was sent after two weeks. To increase the number of respondents, the questionnaire was also distributed via LinkedIn.

Analyses

Factor analysis. To analyze the construct validity of the competency framework, factor analysis was conducted. First, the conditions were checked for factor analysis. No extreme outliers were found. Univariate and multivariate normality was checked using Q-Q plots, histograms, and the Kolmogorov-Smirnov and Shapiro Wilk tests. Almost all items showed significant normality tests, indicating non-normally distributed variables. Furthermore, there is discussion about the minimum sample size for a factor analysis (MacCallum, Widaman, Zhang, & Hong, 1999). Often, the minimum sample is set at 100, or the cases to factors ratio is set at 10. With 76 respondents, this study's sample is not sufficient. However, factor analysis is still used for explorative purpose. Several rules of thumb were used for the creation of variables. Attention was paid to the Kaiser-Meyer-Olkin (KMO) (>.6), a minimum of three items per factor, and the level of communalities (<.5). Cross-loadings can be retained or dropped depending on the interpretability (Yong & Pearce, 2013). Reliability of the scales were evaluated with Cronbach's α (>.7). Names were given to the factors in deliberation with organization X.

Index. The data and technology (DT) domain consisted of general and specific competencies, where the latter refer to specific programs or languages. These specific items were divided into six subdomains which were labeled as databases, analytical modeling, programming languages, big data, general data tools, and business intelligence. For example, analytical modeling included SAS, SPSS, and Stata. A different operationalization approach may be necessary, because being expert in one or a few programs or languages can be more relevant for one's performance than being a beginner in all of them, and different programs can be each other's substitutes. It is therefore questionable whether the items should correlate. Consequently, these items may be better regarded as indexes rather than scales (Streiner, 2003). With scales there are many possible items that tap the construct, and only a sample of them is needed and all items are correlated. However, with indexes the items do not have to correlate, because unrelated items can influence the same latent construct (Streiner, 2003). For indexes it is not useful to perform reliability analysis. Therefore, it was first checked whether there was coherence among the items. Different approaches to operationalization of the variables were tested. First, PAF was performed on the original operationalization of the items. Second, the scales were recoded into three values low, medium, and high. Third, the items were combined similar to the structure in the questionnaire and maximum values were extracted to create new variables. For example, the maximum values from SAS, SPSS, and Stata were computed into a new variable.

Manova. Multivariate multiple regression (MANOVA in SPSS) was used to test the predictive validity of the competencies for performance. Before conducting the analyses, the necessary assumptions were tested. Linearity was assessed through scatterplots and histograms, and Q-Q plots showed to what extent the variable were normally distributed. To check for homoscedasticity, residuals were plotted against predicted values. Finally, for the regression multicollinearity was checked through the correlation matrix (<.08), and variation inflation factor (VIF<10).

Independent sample t-test. Furthermore, independent sample t-tests were executed to identify differences between certain levels of performance. The measurement instrument was evaluated in the three dimensions to indicate whether some competency domains may have better predictive validity than others.

Results Study 2

The outline of results starts with the construction of the competency scales. Thereafter, short scales are created for the competencies and correlated with performance. Finally, the relationship between the competencies and performance dimensions will be tested.

RQ 3&4. The Scales of the Competencies

The competency items were subjected to EFA using Principal Axis Factoring, because the underlying structure was unknown, and EFA allows to explore the latent factors and to place items into meaningful categories (Yong & Pearce, 2013). PAF was unable to explore the underlying structure for all competency items together. Therefore, separate factor analyses were performed.

1. Data and technology. This data and technology (DT) domain consisted of general and specific competencies, where the latter refer to specific programs or languages. First, the general competencies will be discussed, followed by the specific ones. The general DT subdomain contained 9 items. Factor analysis was performed with Oblimin rotation. Both Kaiser criterion and scree plot indicated 2 components with eigenvalue >1 that explained 75.51% of the variance. Two new variables 'programming' and 'data collection & quality' were computed and reliability was good (Cronbach's α .930 and .860 respectively). Results from factor analysis can be found in Table 8.

Table 8 PAF results with Oblimin rotation for general data & technology competencies

Item#	Competencies	Programming	Data collection & quality	Communality
DT1	Programming the right results	1.004		.836
DT3	Collecting information from diverse datasets	.862		.814
DT4	Aggregating/grouping datasets to a higher level	.817		.841
DT2	Programming in an efficient and effective way (technical smart)	.753		.672
DT6	Creating a data and analytics environment in which customer data is available for analyses		.861	.728
DT7	Extracting and using new data sources		.776	.600
DT8	Detecting and solving problems with data quality		.750	.613
DT9	Visualizing results in a simple and concise manner in a dashboard		.713	.613
	Eigenvalue	5.312	1.484	
	% of total variance	59.02	16.49	

Note. Factor loadings <.3 are suppressed, KMO=.858

Different forms of operationalization for the specific items were tested. None of them correlated well enough to result in reliable scales. As expected, this study could not detect coherence between these specific variables and therefore it was more acceptable to treat them as indexes (Streiner, 2003). Therefore, a new variable ‘program index’ was computed counting the amount of high scores (6 and 7) for all data & technology specific variables. This may be more meaningful than analyzing all items separately, since it may not matter what sort of programs or languages data scientists use, as long as they are good in using them. A remark from one respondent supports this approach: “I don’t find all competencies above essential for a good data scientist. With a 7 on one of the programming languages you can go a long way.”

2. Analysis methods and techniques. The domain analysis methods and techniques (AMT) also consisted of both general and specific items. The general competency domain contained seven items. Results from PAF analysis showed a one factor solution. The name AMT general therefore remained the same. The item ‘Producing insights that are new and/or unexpected to the customer’ was deleted in the reliability analysis, since it lowered the reliability. Finally, the scale showed a high Cronbach’s alpha (α=.958). Results from the factor analysis can be found in Table 9.

Table 9 PAF results with no rotation for general analysis methods and techniques competencies

Item#	Competencies	AMT-general	Commu-nality
AMT17	Obtaining the meaning from large data sets	.944	.892
AMT18	Making relevant connections	.931	.866
AMT19	Translating the business question to the right analytical method	.900	.810
AMT20	Applying analytical methods to get the right results	.886	.786
AMT21	Interpreting results correctly	.870	.757
AMT23	Producing insights that are new and/or unexpected to the customer	.784	.614
	Eigenvalue	4.928	
	% of variance	82.14	

Note. Factor loadings <.3 are suppressed, KMO=.903

In the questionnaire the specific items were categorized into three subdomains statistics, machine learning, and other analyses. Unlike for data & technology, for these items it was expected that they did correlate, because it is more likely that data scientists can perform a variety of analyses. Factor analysis (PAF) was performed on all specific AMT items and Oblimin rotation was used since the factors correlated. The two factor solution explained 77.15% of the variance. As expected, the specific items from the AMT domain correlated with each other and these variables will thus be treated as scales and not as indexes. One serious cross loading (‘ensemble, e.g. random forest’) occurred and this item was excluded from further analysis. In reliability analysis ‘time series analysis’ and ‘cohort analysis’ were removed because they lowered reliability. Two new variables ‘basic analytics’ and ‘advanced modeling’ were created. Internal consistencies were high (α=.972) and good (α=.886) respectively. Results from the factor analysis can be found in Table 10.

.

Table 10 PAF results with Oblimin rotation for specific analysis methods and technoqies competencies.

Item#	Competencies	Basic analytics	Advanced modeling	Commu-nalities
AMT24f	Logistic regression	1.005		.893
AMT24e	Linear regression	.975		.880
AMT24c	Clustering analysis	.906		.847
AMT24g	Decision tree/Chaid analysis	.889		.875
AMT24i	Regression model optimization	.889		.844
AMT24d	Factor analysis	.888		.790
AMT24b	Profile analysis	.862		.742
AMT24a	Correspondence analysis (cross table)	.835		.674
AMT24h	Coherence test	.784		.723
AMT24j	Time series analysis	.711		.516
AMT26b	Cohort analysis	.708		.766
AMT25e	Regularization		.916	.774
AMT25f	Instance based		.910	.500
AMT25a	Associated Rule		.809	.681
AMT25d	Neural Networks		.727	.720
AMT25b	Ensemble (e.g. random forest)	.434	.455	.623
	Eigenvalue	10.370	1.974	
	% of variance	64.81	12.34	

Note. Factor loadings <.3 are suppressed, KMO=.895.

3. Impact & advisory and business domain expertise. The impact and advisory (IA) domain consisted of seven subdomains: leading role in the field of data & analytics, structure presentation, personal impact, planning and organizing, team work, results & goal orientation, and customer orientation. The business domain expertise (BDE) domain had four subdomains: business knowledge, analytical applications, creating impact on business goals, and implementing initiatives. The dimensions IA and BDE appeared to be social and business related competencies which may overlap. Therefore, PAF was performed with all items from IA and BDE together. A four factor solution explained 76.69% of the variance. Two cross loadings were not a big problem because there was a difference of approximately .3 between the loadings and interpretability was good. New variables 'Planning & interaction', 'Business knowledge', 'Analytical applications', and 'Presenting results' were created. Planning & interaction is a combination of items from the subdimensions planning & organizing, team work, results & goal orientation, customer communication, and implementing initiatives. Presenting results is a merger of items from former subdimensions 'structured presentation' and 'personal impact'. Reliability scores were high. Cronbach's α was .966 (planning & interaction), .955 (business knowledge), .919 (analytical applications), and .975 (presenting results). Table 11 presents the results from factor analysis.

Table 11 PAF results with Oblimin rotation for items from impact & advisory and business domain expertise competencies.

Item#	Competencies	Planning interaction	Business knowledge	Analytical applications	Presenting results	Commu-nalities
IA57	Empathizing with customers and responding to their needs	.923				.794
IA42	Creating conditions for performing analyses	.843				.646
IA43	Creating clear agreements in consultation with colleagues/customers	.793				.763
IA56	Formulating hypotheses concisely	.773		.306		.723
IA55	Understanding the underlying goals and/or needs with an (analytical) question	.763				.809
IA44	Prioritize shared goals above personal goals	.710				.667
IA48	Deliver impactful output with the appropriate standards of time, quality, and costs	.706				.735
IA41	Differentiating between major and minor issues	.665				.687
IA52	Active listening, summarizing and picturing the input from others	.640				.694
IA47	Providing visible results within agreed deadlines	.626			-.334	.694
IA50	Composing a thorough implementation plan so that results can be obtained	.561				.566
IA40	Planning and organizing activities in a structured way	.558				.590
BDE79	Formulating concrete improvements based on results	.528				.740
IA58	Building and maintaining good relations with stakeholders	.471				.718
BDE66	Knowledge of important stakeholders		.894			.846
BDE60	Knowledge of the business functions/units		.885			.780
BDE61	Knowledge of the business strategy and vision		.883			.835
BDE62	Knowledge of propositions and products		.846			.779
BDE65	Knowledge of the competitive position		.833			.653
BDE63	Knowledge of current problems and opportunities		.823			.807
BDE64	Knowledge of the industry		.816			.728
BDE73	Making a pricing sensitivity model			.760		.689
BDE70	Making a forecasting model			.742		.638
BDE69	Making ROI calculations			.700		.636
BDE68	Estimating the (expected) customer value			.610		.732
BDE67	Providing suiTable analytical solutions			.592		.708
BDE72	Making customer journey insightful			.567		.492
BDE74	Thinking of broad applicable initiatives			.474		.741
IA35	Presenting the storyline in a convincing way to realize the desired goal				-.882	.914
IA33	Translating the message to a structured storyline				-.864	.865
IA34	Visualizing results from analyses in a way that enhances the storyline				-.798	.806
IA32	Translating insights from analyses to a relevant message				-.779	.879
IA37	Presenting in a confident and convincing way				-.723	.818

Table 11 continued

Item#	Competencies	Planning interaction	Business knowledge	Analytical applications	Presenting results	Commu-nalities
IA36	Translating the output in to impact that is relevant for the business				-.711	.902
IA39	Making a strong, professional impression on others				-.637	.814
IA38	‘Selling’ the analysis/project with strong arguments				-.632	.758
IA31	Steering people to deliver impactful output				-.530	.711
	Eigenvalue	22.245	2.698	2.108	1.325	
	% of variance	60.12	7.29	5.70	3.58	

Note. Factor loadings <.3 are suppressed, KMO=.893.

The previous section provide an answer to research questions 3 and 4. The competencies that resulted from the analyses are programming, data collection & quality, specific program indexes, analysis methods & techniques (AMT) general, basic analytics, advanced modeling, planning & interaction, business knowledge, analytical applications, and presenting results.

RQ 5. Scale Reduction

All the scales that were created showed high internal consistencies and contained between four and fourteen items. To reduce the amount of items, the highest loading items were used to create shorter, but still reliable scales. The scales were shortened with 36 items. The short scales are presented in Table 12. A few items from basic analytics were combined into a new item as ‘regression techniques’, and another few were combined into a new item ‘explorative techniques’. Worth noting is the competency planning & interaction, where seven out of fourteen items were removed. The mean score and reliability remained largely the same. Also, four out of seven items from business knowledge were deleted, since the questions were similar to each other. As the original scale contained 108 items, the short scale is 64 items shorter.

Table 12 Descriptives, reliabilities and number of items for long and short scales of the competencies.

	Scales from analysis			Short scales		
Competencies	Mean[2](SD)	α[1]	#items	Mean(SD)	α[1]	#items
Programming	4.92 (1.31)	.930	4	5.16(1.34)	.920	3
Data collection & quality	4.52(1.36)	.860	4	4.43(1.38)	.852	3
Program index	1.84[3](1.62)	-	17	-	-	6
AMT general	4.96(1.27)	.958	5	4.94(1.31)	.957	3
Basic analytics	4.23(1.67)	.972	9	4.17(1.65)	.961	5
Advanced modeling	1.84(1.15)	.886	4	1.69(1.14)	.870	3
Planning & interaction	4.68(1.05)	.966	14	4.72(1.04)	.946	7
Business knowledge	4.60(1.32)	.955	7	4.63(1.27)	.937	3
Analytical applications	4.55(1.21)	.919	7	3.85(1.36)	.905	6
Presenting results	3.62(1.34)	.975	9	4.72(1.34)	.965	5
Total			80			44

Note. [1]Cronbach's alpha. [2]1 (no experience) to 7 (expert). [3]Amount of 6&7 scores.

RQ 6. Predictive Validity of Competencies on Performance

Correlations. Correlations were computed to examine the relationships between the performance dimension, the short scale competencies and control variables. Among the performance dimensions, task performance and OCB positively correlated (r =.455, p <.001). CWB did not correlate with other performance variables, nor with any of the competencies. All competencies showed positive and significant correlations with OCB. For example, data collection & quality (r =.451, p <.001) and planning & interaction (r =.431, p <.001) were positively related to OCB. Task performance showed positive correlations with planning & interaction (r =.254, p =.028). Almost all competencies correlated with each other. None of the control variables correlated with the performance dimensions. They did show correlations with several competencies. For example, the positive correlation between age and business knowledge (r =.320, p =.005) indicated that older data scientists have more knowledge about the business. Also, education showed a positive relationship with basic analytics (r =.360, p =.001), presenting results (r =.320, p =.005) and analytical applications (r =.270, p =.019). Finally, results showed that men scored higher on analytical applications than women (r =.262, p =.023). Correlations and descriptives of the short scale competencies, performance dimensions, and control variables are presented in Table 13.

Table 13 Correlations and descriptives of performance dimensions, short scale competencies, and control variables.

		Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	TP	3.50	0.62	(.705)[2]														
2	OCB	3.67	0.65	.455**	(.824)													
3	CWB	4.05	0.67	.029	.061	(.840)												
4	Programming[3]	5.16	1.34	.140	.305**	-.070	(.920)											
5	Data collection & quality	4.43	1.38	.011	.451**	.033	.525**	(.852)										
6	Program index	1.84	1.62	.020	.252*	-.193	.684**	.442**										
7	AMT general	4.94	1.31	.172	.297**	-.072	.708**	.570**	.619**	(.957)								
8	Basic analytics	4.18	1.65	.048	.341**	-.188	.720**	.397**	.679**	.741**	(.961)							
9	Advanced modeling	1.69	1.14	.028	.277*	-.115	.406**	.235*	.471**	.403**	.484**	(.870)						
10	Planning & interaction	4.72	1.04	.254*	.431**	.093	.574**	.593**	.455**	.710**	.464**	.332**	(.946)					
11	Business knowledge	4.63	1.27	.043	.284*	.091	.246*	.391**	.222	.520**	.330**	.145	.658**	(.937)				
12	Analytical applications	3.85	1.36	.057	.424**	-.055	.536**	.470**	.538**	.694**	.723**	.354**	.657**	.578**	(.905)			
13	Presenting results	4.72	1.34	.107	.292*	-.002	.580**	.492**	.521**	.783**	.668**	.421**	.725**	.554**	.711**	(.965)		
14	Gender[4]	0.58	0.50	-.114	.050	-.154	.084	.151	.132	.178	.060	.084	.221	.166	.262*	.185		
15	Age[5]	2.89	1.41	-.129	-.020	.186	-.071	.223	-.048	-.016	-.193	-.049	.203	.320**	.075	-.032	.221	
16	Education[6]	3.62	0.94	-.004	-.100	-.068	.015	-.068	.171	.191	.360**	.163	.052	-.007	.270*	.320**	-.120	-.283*

Note. ** p<0.01, * p<0.05. [1]Standard deviation. [2]Cronbach's alpha. [3]4-10 are short scales. [4]0=female, 1=male. [5]From 1 (<26 years) to 6 (46-50 years). [6]From 1 (high school/vocational training) to 5 (PhD or other advanced degree).

Manova. In order to test for a relationship between the competencies and performance dimensions, multivariate regression analysis (Manova in SPSS) was conducted with three dependent variables, the performance dimensions, and ten competency variables as independent variables. Results did not provide significant outcomes. For example, multivariate tests showed one statistically significant difference for the competencies on the combined dependent performance dimensions, which was for data collection & quality: $F(3, 74)=6.495$, $p=.001$; Wilks' Lambda=.758. When analyzing the results for the performance dimensions separately, the only significant difference ($p<.5$) was for OCB: $F(1, 74)=6.925$, $p=.011$. However, when using a Bonferroni adjusted alpha level (Pallant, 2005) of .005 to reduce the chance for a type 1 error, this result was not statistically significant. All other competencies did not show significant differences for the combined performance dimensions. For brevity, only the results for data collection & quality are presented in Table14. This analysis could thus not provide sound evidence for predictive validity of competencies on performance.

Table 14 Multivariate regression (Manova) for IV data collection and quality on the IV performance dimensions

IV	Wilks' Lambda	Df	F	Sig	DV	Df	F	Sig
Data collection & quality	.758	3	6.495	.001	TP	1	2.109	.151
					OCB	1	6.708	.012
					CWB	1	.062	.803

Note. N=76.

Independent sample t-tests. To further investigate the relationships between the competencies and performance, independent sample t-tests were conducted to analyze whether competencies were able to distinguish the top performers from average performers. To identify the top performers, approximately 10% of the highest performers of each performance dimension was selected. To select a group of average performers, approximately 25% who scored around the 50% of all respondents was selected for each performance dimension. Since different respondents can have the same performance scores, it was not always possible to extract an optimal number of respondents. For example, for CWB a good balance of respondents below and above the 50% average in CWB scores was preferred over the percentage of 25% of respondents, resulting in a higher percentage of 34.2%. An overview of the selection can be found in Table 15.

Table 15 Overview of N, scores, and percentages for the top and average performers

	Top performers (N)	Score	%	Average performers (N)	Score	%
TP	11	>=4.25	14.4	21	3.25-3.50	27.6
OCB	7	>=4.57	10.4	11	3.50-3.75	19.7
CWB	6	5	7.9	26	4.0-4.4	34.2

Task performance. First, independent sample t-tests were performed to analyze the differences in competencies for average and top performers in task performance. No significant differences were found. There is thus no evidence that proficiency levels in competencies are statistically different between top and average performers. Results are presented in Table 16.

Table 16 Independent sample t-tests for competencies with average and top performers in task performance.

	TP[4] average[5]			TP top[6]			Difference		
	M[1]	SD[2]	N[3]	M	SD	N	t	df	Sig.
Programming	5.43	1.28	21	5.55	0.95	11	0.266	30	.792
Data collection & quality	4.52	1.40	21	4.24	1.56	11	-0.521	30	.606
Program index	2.19	1.40	21	2.0	1.00	11	-0.399	30	.692
AMT general	5.19	1.28	21	5.42	1.25	11	0.494	30	.625
Basic analytics	4.90	1.60	21	4.12	1.45	11	-1.350	30	.187
Advanced modeling	1.79	1.13	21	1.73	0.83	11	-0.172	30	.865
Planning & interaction	4.68	1.15	21	5.05	0.95	11	0.917	30	.366
Business knowledge	4.62	1.52	21	4.55	1.06	11	-0.143	30	.887
Analytical applications	4.13	1.42	21	3.83	1.23	11	-0.579	30	.567
Presenting results	4.75	1.50	21	5.22	1.01	11	1.044	27.8	.305

Note. [1]Mean. [2]Standard deviation. [3]Number of respondents. [4]Task performance. [5]TP=3.25-3.50. [6]TP>= 4.25.

OCB. The same analysis was performed with OCB. All differences were positive except for 'basic analytics'. For planning & interaction the mean score was significantly larger for OCB top performers than for average performers (Mdif=.93; t=.2.226, *p*=.038). Additionally, top performers scored significantly higher on business knowledge than average performers did (Mdif=1.37; t=2.624, *p*=.016). From these results it can be concluded that the competencies planning & interaction and business knowledge are able to distinguish among top performing and average performing data scientists in OCB. Results are presented in Table 17.

Table 17 Independent sample t-tests for competencies with average and top performers in OCB.

	OCB[4] average[5]			OCB top[6]			Difference		
	M[1]	SD[2]	N[3]	M	SD	N	t	df	Sig.
Programming	5.29	1.35	15	5.50	1.27	8	.364	21	.720
Data collection & quality	4.47	1.07	15	5.33	1.38	8	1.676	21	.109
Program index	1.87	1.55	15	2.88	2.64	8	1.161	21	.259
AMT general	5.20	1.26	15	5.25	1.15	8	0.093	21	.927
Basic analytics	4.62	1.71	15	4.53	1.90	8	-0.117	21	.908
Advanced modeling	2.02	1.36	15	2.58	2.07	8	0.786	21	.441
Planning & interaction	4.62	0.97	15	5.55	0.78	7	2.226	20	.038
Business knowledge	3.91	1.20	15	5.29	0.99	7	2.624	20	.016
Analytical applications	3.83	1.44	15	4.74	1.07	7	1.480	20	.154
Presenting results	4.52	1.30	15	4.97	1.43	7	0.729	20	.475

Note. [1]Mean. [2]Standard deviation. [3]Number of respondents. [4]Organizational citizenship behavior. [5]OCB=3.50-3.75. [6]OCB>= 4.57.

CWB. Finally, the analysis with CWP showed that no significant differences in competencies exist between the CWB performance groups. Results can be found in Table 18.

Table 18 Independent sample t-tests for competencies with average and top performers in CWB.

	CWB[4] average[5]			CWB top[6]			Difference		
	M[1]	SD[2]	N[3]	M	SD	N	t	df	Sig.
Programming	5.15	1.27	26	5.61	1.10	6	0.814	30	.422
Data collection & quality	4.37	1.18	26	5.00	1.41	6	1.139	30	.264
Program index	1.77	1.95	26	1.50	1.52	6	-0.316	30	.754
AMT general	4.88	1.08	26	5.33	1.63	6	0.833	30	.411
Basic analytics	3.89	1.63	26	4.42	1.19	6	0.752	30	.458
Advanced modeling	1.58	1.25	26	1.67	0.52	6	0.171	30	.865
Planning & interaction	4.77	0.89	26	4.79	1.24	6	.0380	30	.970
Business knowledge	4.91	1.11	26	4.22	1.26	6	-1.341	30	.190
Analytical applications	3.86	1.24	26	3.67	1.48	6	-0.331	30	.743
Presenting results	4.66	1.30	26	4.57	0.95	6	-0.165	30	.870

Note. [1]Mean. [2]Standard deviation. [3]Number of respondents. [4]Counterproductive work behavior. [5]CWB=4.0-4.4. [6]CWB= 5.0.

In short, none of the competencies were able to distinguish top from average performers in task performance and CWB. However, two competencies, business knowledge and planning & interaction, accounted for a difference in mean scores between top and average performers in OCB. There is thus an indication for predictive validity of these competencies, as they can distinguish top from average performers in OCB.

RQ 3. Missing Competencies

Twenty-six respondents provided useful feedback on the question whether items were missing which are evident for a good data scientist. Visualization was mentioned four times, Tableau eleven times, and PowerBI five times. This indicates that visualization is a competency that should be further explored. Visualization will therefore be added as a program index in the new questionnaire with Tableau and PowerBI as optional programs. Additionally, one respondent suggested an item that included having curiosity, playing devil's advocate and questioning the data and analyses. An item with this content will be suggested as well for the new questionnaire.

Development of the Competency Domains

By revising the competency framework, the competency domains have been changed. Table 19 presents an overview of the development of the competencies in this study. The new questionnaire with short scales and adaptions from feedback is presented in Appendix 5.

Table 19 Development of the competency domains from the original, revised, and new framework.

General domain	Framework study 1	Revised framework study 2	New framework
Data and Technology (DT)	General D&T	General D&T	Programming
	Basic programming languages		Architecture & quality
	Advanced programming languages	Specific D&T	Specific program index
Analysis Methods and Techniques (AMT)	General AMT	General AMT	General AMT
	Basic analytics - explorative	Statistics	Basic analytics
	Basic analytics - statistics	Machine learning	Advanced analytics
	Advanced analytics – machine learning	Other analyses	
Impact and Advisory (IA)	Structured presentation	Structured presentation	Presenting results
	Personal impact	Personal impact	
	Planning and organizing	Planning and organizing	
	Results & goal orientation	Results & goal orientation	
	Communication skills	Customer communication	Planning & interaction
	Team work	Team work	
	Leading role in the field of data & analytics	Leading role in the field of data & analytics	
	Customer orientation		
Business Domain Expertise (BDE)	Knowledge of the business	Knowledge of the business	Business Knowledge
	Analytical applications	Analytical applications	Analytical applications
	Creating impact on business goals	Creating impact on business goals	
	Implementing initiatives	Implementing initiatives	

Discussion

Findings

Organization X had developed a competency model for data scientists with four domains, which were data & technology (DT), analysis methods & techniques (AMT), impact & advisory (IA), and business domain expertise (BDE). This research sought to improve and validate this competency model in order to predict the performance of data scientists. This research consisted of two studies. Study 1 was mainly concerned with the improvement of the competency framework and the questionnaire, and some explorative analyses. Study 2 was also explorative and attempted to create reliable scales, define the competencies, shorten the scales, and finally validate the revised measurement instrument. 21 data scientists from organization

X participated in study 1 by completing a questionnaire, and in study 2 there were 76 respondents from other organizations who filled in the adapted questionnaire. The findings followed by possible explanations for the results are outlined below.

Study 1. ***RQ 1 performance measurements***. To assess the convergent validity, the performance measurements from the self-reports, the supervisor ratings, and the official ratings were correlated. In study 1 these correlations were not found, which does not provide evidence that these measures actually measure the same performance criterion. Although they did not correlate, supervisor ratings did also not significantly differentiate from the self-reports. Explanations for the absent correlations may the low sample size, biases that arouse with self-reports and ratings by others, or that the constructs are actually different from one another. The average performance scores were all high when compared to the norm scores for white collar workers (Koopmans, 2015). It is possible that the data scientists actually perform high. Another possibility is that biases such as overestimation and leniency bias caused data scientists and supervisors to rate performance higher than the true performance (Bol, 2011).

RQ 2 GCA and performance. It was expected that a higher GCA would be related to higher performance. However, no significant differences in performance were found for people with a higher intelligence. This may be due to low discriminability of the GCA scores, since all employees are above academic intelligent. Also, the cut point of 270 put two people with a score of 265 and 269 in the ‘lower’ group, whereas three people with slightly higher scores of 271 and 273 were put in the ‘higher group’. The positive relationship between GCA and performance may still be there, but it could not be detected with this selected group. It is also possible that because data scientists generally have high intelligence, other factors such as competencies are more relevant (Spencer & Spencer, 2008).

RQ 3&4 competencies and items. Face validity was assessed through feedback from data scientists. The competency framework was adjusted and some missing competencies were added. Amongst other things, it was found that making a connection between analyses and the implications for the business was important for data scientists.

RQ 6 predictive validity of competencies on performance. There were indications for a positive relationship between some competencies and task performance and OCB, but not with CWB. Because of the low sample size, no factor and reliability analyses were performed and therefore the underlying structure and internal consistencies could not be analyzed. These results were only for explorative purpose and the relationships will be discussed more in detail for study 2.

Study 2. ***RQ 3&4 the scales of the competencies.*** Based on factor and reliability analyses, ten competency variables were computed in study 2. When comparing the new competency subdomains with the original ones, several differences can be noticed. First, the general data & technology competency was split into two competencies programming and data collection & quality. Competencies in programming can thus be seen as distinct from competencies in data collection and quality. The specific items from data & technology were computed as indexes, because the items did not correlate and could not be placed in a scale. This approach seems plausible since being good in one specific program or language is more likely to result in high performance than being a beginner in a variety of competencies. Moreover, from the respondents' feedback it became clear that not all data scientists use the same programs and that the options in the questionnaire were not exhaustive. Different programs may be complementary or substitutes. Second, the general analysis methods & techniques variable remained the same. The specific items were translated into two subdomains basic analytics and advanced modeling. Caution with this distinction is necessary, because some analyses can be both basic and advanced, and this may depend on one's job or project. Third, the subdomains from impact & advisory were changed to presenting results, interaction with others, and planning & results. The latter two variables were not supported by factor analysis, as the analysis showed that they belonged in one factor. These variables therefore need extra attention when interpreting results. Finally, business knowledge and analytical applications remained the same. The subdimensions 'creating impact on business goals' and 'implementing initiatives' from the original competency model were not found through factor analyses, since most of these items had been deleted due to low communality.

RQ 5 short scales. Short scales were created to abbreviate the questionnaire and to lower problems of survey fatigue. The items were removed based on their factor loadings, the reliability scores, and the final correlations with performance. Because the minimum amount of respondents had not been met for factor analysis, caution is necessary with the scale reduction based on factor loadings.

RQ 6 predictive validity of competencies. Ten competencies for data scientists were created and all of them positively correlated with OCB, only planning & interaction correlated with task performance, and none of the competencies correlated with CWB. An indication for predictive validity was found for business knowledge and planning & interaction on OCB. When comparing these findings with previous literature, business knowledge is in line with, for example, Davenport (2012), who argued that business knowledge is essential for data scientists. The competency planning & interaction seems to fit best with the communication, as proposed by, for example, Kolassa (2016). No relationship was found for the other competencies. One possible explanation is that these competencies are simply not related to the performance of data scientists. Furthermore, other important predictors of performance, such as engagement (Rich et al., 2010) and well-being (Peccei et al., 2013), were not controlled for in this study. When comparing these results with previous literature, the absence of evidence for predictive validity of technical and analytical competencies is highly unexpected. A more likely reason for the findings is that the items and dimensions did not adequately reflect the actual competencies of data scientists.

To understand the different results for the three performance dimensions, the scales are further inspected. First, the task performance scale was mainly about time management and results. The correlation with planning & interaction is therefore straightforward, since this variable includes items about planning and results. The other competencies may have no influence on task performance. For example, having a lot of business knowledge may not be related to the ability to manage time. Furthermore, it is possible that task performance is influenced to a greater extent by other predictors, such as engagement (Rich et al., 2010). Engaged employees work more intensely for longer periods of time and are more focused on responsibilities and results (Rich et al., 2010). Second, the OCB scale mainly included items about being active, seeking challenges, taking responsibility, and developing skills and knowledge. Most likely, the competencies correlate with OCB, because the higher the competency proficiency, the more likely data scientists develop skills and knowledge, look for challenges, and take extra responsibilities. Causality cannot be inferred. In fact, it also appears logical that when people develop themselves, seek challenges and take responsibilities, their competencies will grow. One possible theoretical explanations is that this relationship is influenced by other factors, such as a learning organizational culture (Jo & Joo, 2011). This culture is characterized by, amongst others, continuous learning, empowerment, and knowledge sharing (Jo & Joo, 2011). Maybe data scientists generally work in learning cultures. Finally, in studies 1 and 2 no correlations were found between competencies and CWB. Maybe the competencies do not relate to this part of performance, because being good at, for example, programming and resenting results may not influence the perceived negative aspects of work and vice versa. As discussed earlier, the used CWB scale does not accurately reflect the construct of CWB. Counterproductive work behavior includes detrimental behavior such as sabotage, theft, and absenteeism, but this was not reflected in the items. The CWB items were rather concerned with the negative aspects of work and complaining. Next to that, there may be other factors that may influence CWB but which were not measured in this research, for example 'self-control'(Marcus & Schuler, 2004). People are high in self-control when they tend to avoid behavior whose long-term costs are higher than short-term advantages. According to Angrave et al. (2016) self-control explained a great amount of CWB's variance and may therefore be an important variable. In the end it is questionable whether this performance dimension is relevant at all when evaluating and predicting the performance of data scientists. A relationship between competencies and OCB may be more relevant, since CWB is extreme behavior and may not occur frequently. Also, employees high in OCB are more likely to outperform others and thus distinguish the top performers (Emami, Alizadeh, Nazari, & Darvishi, 2012).

Conclusion

The main purpose of this research was to improve and validate a competency model for data scientists in order to predict their performance. No convergent validity was found for the performance measurements and no relationship was found between general cognitive ability and the performance of data scientists. The original competency framework was revised based on feedback from data scientists. This questionnaire was distributed among a larger sample of data scientists. Based on the collected data the competencies were grouped together into ten competencies with reliable scales and compared with the original framework. Whereas some competencies remained largely the same, other subdomains deviated from the original framework. This mainly resulted in fewer subdomains, since some scales were merged or excluded. In addition, shorter but still reliable scales were suggested. This study was able to provide evidence for indications of predictive validity of the competencies business knowledge and planning & interaction on OCB. Grounded evidence of predictive validity could not be provided, since the results were found in correlations and t-tests, rather than predictive analyses. Furthermore, the competencies were mostly related to OCB, and secondly to task performance. CWB did not relate to the competencies at all. This study contributed to the validation of a competency model for a complex and rare profile, which is relevant for both practice and academics.

Limitations and Future Research

Nine important limitations will be discussed below and thereupon suggestions for future research are given.

First of all, there are issues with competencies. Competencies are hard to define and the difference with performance is not clear cut, especially when performance and competencies are both defined as behavior (Markus et al., 2005) which is the case in this study. Competencies can also be included in the performance domain (Bartram, 2005) which presumably occurred with ‘planning & results’ and ‘task performance’. Caution is necessary as this leads to circularity in reasoning. However, fully disregarding competencies may not be a solution, as they can still be seen as an enabler for performance (Kurz & Bartram, 2002). Further investigation about competencies and performance is essential to expose the true relationship.

Next to that, using self-reports to measure competencies and performance is subjective and comes with many possible biases, such as overestimation (Viswesvaran, 2001). The answers given in the questionnaire may represent a twisted truth. To more correctly test the level of competencies and performance, 360-degree feedback could be used lower the rater bias and to better measure performance (Atwater et al., 2002). Also, trained raters could evaluate competencies as performed in practice.

Additionally, other important predictors of performance, such as GCA and personality (Kanfer & Kantrowitz, 2002) were not included in the second study. It was therefore not possible to test these relationships with the performance dimensions in presence of the competencies. This would have been useful to analyze discriminability and the added value of measuring competencies.

Another, more specific, limitation encompasses the incomplete program indexes that were used in the questionnaires. Some data scientists were therefore not able to rate their proficiency level in all relevant fields, for example in visualization programs. The questionnaire should be adapted to overcome this issue. Since it is not feasible to include all possible options, another solution is required. This study suggests that when it is clear which specific technical competencies data scientists needs for their jobs, these should be included in the questionnaire. However, when this is not the case, as in this research, it may be a better option to ask the data scientists in an open question which programs they use for work and subsequently ask their proficiency in those competencies. This approach is likely to make the questionnaire valuable for many different data scientists.

Furthermore, another limitation concerns the feedback and improvements of the competency framework. Although feedback was collected among data scientists, no panel discussion about the competency model and questionnaire was held. This could have enhanced understanding and consensus about adjustments and a comprehensive discussion could have led to a more in depth analysis of the competency framework and the questionnaire. Moreover, the respondents filled in the questionnaire which contained many items, and were also asked to provide feedback. It is questionable whether they had spent enough time and effort to genuinely evaluate the domains and the items. It is therefore possible that some changes are still necessary but have not been detected. Consequently, it is advised to carefully check the new scales.

Also, this study did not consider different data scientist roles as researchers did in Davenport et al. (2001). They argued that the analytical talent needs skills in all five skillsets that they proposed, but to a different extent. This study did not account for those particular differences but rather implicitly assumed that all respondents need all competencies equally for good performance regardless of their role. Further research could study those data scientists' roles more in depth. Furthermore, the competencies largely correlate with each other, making it plausible that higher order factors could be extracted. Further research could examine these higher order factors and compare these with existing literature.

A more statistical drawback involves the small sample size. In study 1, only a few performance ratings and GCA scores could be merged, leading to doubtful results. In study 2, the small sample consisted of both Dutch and international people, which may suffer from different interpretations and cultural differences. On top of that, the small sample size was officially not sufficient to execute factor analysis, as a larger sampling error occurs (MacCallum et al., 1999). Especially low communalities can be influenced by a sampling error. Although items with low communalities were excluded, the small sample can still cause different results. This sampling problem also extends to all other analyses. It is plausible that at least some of the significant correlations or t-test results are found by chance. In a larger sample, analyses may show different outcomes, and it is possible that other dimensions with different items would be found. It is therefore advised to test the measurement instrument on a larger sample of data scientists.

In addition to the previous limitation, other analyses could have been used that are more robust and can partly restore the issue of the small sample size. For example, a simulation with nonparametric bootstrapping could have been performed, where resamples from the sample data are drawn and for every resample the statistic is calculated. A bootstrap confidence interval provides more reliable insights than a single test score. Although the original sample still causes the most variation, the resampling in this method results in better estimates about the true statistic.

Finally, this study was ambitious in obtaining predictive validity of the competency measurement instrument. The study's method actually did not allow to test for predictive validity, since prediction involves testing new instances. Rather, this study developed and improved a model, and searched for indications of predictive validity. True prediction is the next step for further research. Next to that, (k-fold) cross validation could have been used, where the data is split (k times) into a training and test set. First, this provides better results in developing the model (thus in selecting the competency domains), and second, this allows the researcher to estimate the prediction error. In the end, quality could have been improved when splitting the research objectives and first focus on the model itself, as this research attempted to both improve a competency model and find predictive validity in one study.

In addition to the research suggestions based on the limitations, further research could take a total different approach in developing a data scientist profile. For example, by scraping data from Linkedin profiles a different model can be developed about the profile of a data scientist. The main advantage is the availability of big data that can provide large scale insights.

Implications

The outcomes of this research come with both practical and scientific implications. The first practical implication derives from the findings on the underlying structure of the competencies and its relationship with performance. Organization X will be able to further improve their competency framework for data scientists. A better competency framework and corresponding questionnaire will provide more insights about the characteristics of a good performing data scientist and increases the measurement accuracy. This consequently gives more precise information for decision-making on performance improvements. Amongst others, this can include practices concerning the training and development. For example, competencies from planning & interaction maybe a focus area for the training programs of organization X.

The second implication concerns specific improvements in the questionnaire. For example, indexes for the specific data and technology competencies can replace the incomplete scales and they can be adjusted based on the data scientists' requirements. Instead of including a list of programs and languages, only the relevant competencies can be asked. This makes the questionnaire more reliable and applicable to all data scientists. Finally, the short scales can make the measurement instrument more practical and reduces survey fatigue.

Apart from the implications based on this study's results, one could reflect upon using the questionnaire as a measurement instrument in general. Is a questionnaire the best way to evaluate competencies and performance and should organizations use it? In order to answer this question, one needs to be familiar with the alternatives. One alternative is to organize focus groups or to do interviews to collect more in depth insights about competencies and performance. For example, by interviewing multiple stakeholders about the competencies and performance of a person, more information can be gathered. An obvious drawback is the resources that are necessary to conduct such an intensive study, especially when doing this on a large scale. Alternatives for measuring performance could be to analyze promotions or salaries, which have the advantage over questionnaires that the researcher does not need to invest time in designing and collecting questionnaires. However, such measurements come with their own limitations, like criterion contamination and criterion deficiency. It seems that there is no best alternative. A questionnaire may be a balance between these options. It collects more direct information about competencies and performance than the objective measurements, but it is less time consuming that performing focus groups and in depth interviews. Another interesting question to ask is: What does an influential organization as Google do? It is not clear whether they use competencies at all, but the performance measurement system is known. Google uses a 5-point rating scale from ‘needs improvement’ to ‘superb’ to measure employee performance. Peer reviewers provide ratings, and subsequently managers have calibration sessions where rater bias is eliminated as much as possible. Laszlo Bock, former HR chief at Google, argues that this calibration is crucial. Organizations should not only pay attention to the rating itself, but also to a thorough assessment of the results to enhance consistency.

The first scientific implication concerns the contribution to scientific research in HR analytics and specifically in the unexplored competency framework for data scientists. This research offers the opportunity for other researchers to further refine and validate the competency domains that emerged from the factor analyses and thus further develop the competency profile for data scientists. Additionally, since HR needs more data driven practices, this study has contributed to the credibility of HR as a valuable business function and steers HR towards it desired role as a strategic partner (Ulrich, 1997). Furthermore, only few studies have studied the competency framework for data scientists and the small sample size indicates the difficulties of doing research in this field. This study laid some foundations for other researchers to advance research in the competency framework for data scientists and its predictive relationship with performance, and emphasizes the urge for more research on data scientists. After all, it is the sexiest job of the 21st century.

References

Allen, J., & van der Velden, R. (2005). *The role of self-assessment in measuring skills*. Retrieved from https://cris.maastrichtuniversity.nl/portal/files/4047131/guid-3b801703-f49b-46e5-bd15-adfe7eef189e-ASSET1.0

Altink, W., & Verhagen, H. (2002). Assessing potential and future performance. In S. Sonnentag (Ed.), *Psychological Management of Individual Performance* (pp. 179-198). Chichester, UK: John Wiley & Sons.

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: Why HR is set to fail the big data challenge. *Human Resource Management Journal, 26*(1), 1-11. doi:https://doi.org/10.1111/1748-8583.12090

Armstrong, M., & Taylor, S. (2014). *Armstrong's handbook of human resource management practice* (10 ed.). London, England: Kogan Page.

Arthur Jr, W., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of applied psychology, 93*(2), 435-442. doi:http://doi.org/10.1037/0021-9010.93.2.435

Atwater, L. E., Waldman, D. A., & Brett, J. F. (2002). Understanding and optimizing multisource feedback. *Human Resource Management, 41*(2), 193-208. doi:https://doi.org/10.1002/hrm.10031

Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of applied psychology*, 1185-1203. doi:https://doi.org/10.1037/0021-9010.90.6.1185

Bartram, D., Robertson, I. T., & Callinan, M. (2002). Introduction: A framework for examining organizational effectiveness. In D. Bartram, I. T. Robertson, & M. Callinan (Eds.), *Organizational effectiveness: The role of psychology* (pp. 1-10). Chichester, UK: John Wiely & Sons.

Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review, 86*(5), 1549-1575. doi:https://doi.org/10.2308/accr-10099

Boyatzis, R. (2008). Competencies in the 21st century. *Journal of management development, 27*(1), 5-12. doi:https://doi.org/10.1108/02621710810840730

Brogden, H. E., & Taylor, E. K. (1950). The theory and classification of criterion bias. *Educational and psychological measurement, 10*(2), 159-183. doi:https://doi.org/10.1177/001316445001000201

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35-70). San Fransisco, CA: Jossey-Bass.

Cardy, R. L., & Selvarajan, T. (2006). Competencies: Alternative frameworks for competitive advantage. *Business Horizons, 49*(3), 235-245. doi:https://doi.org/10.1016/j.bushor.2005.09.004

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England: Cambridge University Press.

Cascio, W. F., & Boudreau, J. A. (2011). *Investing in people: Financial impact of human resource initiatives*. Upper Saddle River, NJ: Pearson Education.

Cascio, W. F., & Ramos, R. A. (1986). Development and application of a new method for assessing job performance in behavioral/economic terms. *Journal of applied psychology, 71*(1), 20. doi:https://doi.org/10.1037//0021-9010.71.1.20

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of educational psychology, 54*(1), 1. doi:http://dx.doi.org/10.1037/h0046743

Chiang, R. H., Goes, P., & Stohr, E. A. (2012). Business intelligence and analytics education, and program development: A unique opportunity for the information systems discipline. *ACM Transactions on Management Information Systems, 3*(3), 12. doi:https://doi.org/10.1145/2361256.2361257

Conway, D. (2013, March 26). The data science venn diagram. Retrieved from http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Davenport, T. H. (2012). *The human side of big data and high-performance analytics*. Retrieved from http://www.datascienceassn.org/sites/default/files/Human%20Side%20of%20Big%20Data%20and%20High-Performance%20Analytics.pdf

Davenport, T. H., Harris, J. G., de Long, D. W., & Jacobson, A. L. (2001). Data to knowledge to results: Building an analytic capability. *California Management Review, 43*(2), 117-138. doi:https://doi.org/10.2307/41166078

Davenport, T. H., & Patil, D. J. (2012). Data scientist: The sexiest job of the 21st century. *harvard business review, 90*(5), 70-76.

Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J., & Starr, J. M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence, 28*(1), 49-55. doi:https://doi.org/10.1016/S0160-2896(99)00031-8

Debnath, S. C., Lee, B. B., & Tandon, S. (2015). Fifty years and going strong: What makes behaviorally anchored rating scales so perennial as an appraisal method? *International Journal of Business and Social Science, 6*(2).

Debortoli, S., Müller, O., & vom Brocke, J. (2014). Comparing business intelligence and big data skills: A text mining study using job advertisements. *Business & Information Systems Engineering, 6*(5), 289-300. doi:10.1007/s12599-014-0344-2

Edwards, M. R., & Edwards, K. (2016). *Predictive HR analytics: Mastering the HR metric*. London, England: Kogan Page.

Emami, M., Alizadeh, Z., Nazari, K., & Darvishi, S. (2012). Antecedents and consequences of organisational citizenship behaviour (OCB). *Interdisciplinary Journal of Contemporary Research in Business, 3*(9).

Fisher, C. (2016). *LinkedIn unveils the top skills that can get you hired in 2017*. Retrieved from https://blog.linkedin.com/2016/10/20/top-skills-2016-week-of-learning-linkedin

Fitz-enz, J. (2010). *The new HR analytics: Predicting the economic value of your company's human capital investments*. New York, NY: AMACOM.

Flamholtz, E. G. (2012). *Human resource accounting: Advances in concepts, methods and applications*. New York, NY: Springer

Goffin, R. D., Gellatly, I. R., Paunonen, S. V., Jackson, D. N., & Meyer, J. P. (1996). Criterion validation of two approaches to performance appraisal: The behavioral observation scale and the relative percentile method. *Journal of Business and Psychology, 11*(1), 23-33. doi:https://doi.org/10.1007/bf02278252

Harris, H. (2013, September 19). The data products venn diagram. Retrieved from http://www.datacommunitydc.org/blog/2013/09/the-data-products-venn-diagram

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of vocational behavior, 29*(3), 340-362. doi:http://dx.doi.org/10.1016/0001-8791(86)90013-8

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law, 2*(3-4), 447-472. doi:https://doi.org/10.1037//1076-8971.2.3-4.447

Jo, S. J., & Joo, B.-K. (2011). Knowledge sharing: The influences of learning organization culture, organizational commitment, and organizational citizenship behaviors. *Journal of Leadership & Organizational Studies, 18*(3), 353-364. doi:https://doi.org/10.1177/1548051811405208

Kanfer, R., & Kantrowitz, T. M. (2002). Ability and non-ability predictors of job performance. In S. Sonnentag (Ed.), *Psychological Management of Individual Performance* (pp. 27-50). Chichester, UK: John Wiley & Sons.

Kolassa, S. (2016, July 8). The new data scientist venn diagram. *StackExchange*.

Koopmans, L. (2015). Handleiding voor de individuele werkprestatie vragenlijst (IWPV).

Koopmans, L., Bernaards, C. M., Hildebrandt, V. H., de Vet, H. C., & van der Beek, A. J. (2014). Measuring individual work performance: Identifying and selecting indicators. *Work*, *48*(2), 229-238. doi:http://doi.org/10.3233/WOR-131659

Koopmans, L., Bernaards, C. M., Hildebrandt, V. H., Schaufeli, W. B., de Vet, H. C., & van der Beek, A. J. (2011). Conceptual frameworks of individual work performance: A systematic review. *Journal of Occupational and Environmental Medicine, 53*(8), 856-866. doi:https://doi.org/10.1097/jom.0b013e318226a763

Koopmans, L., Bernaards, C. M., Hildebrandt, V. H., Van Buuren, S., Van der Beek, A. J., & De Vet, H. C. (2014). Improving the individual work performance questionnaire using rasch analysis. *Journal of Applied Measurement*, *15*(2), 160-175. doi:http://dx.doi.org/10.1136/oemed-2013-101717.51

Kurz, R., & Bartram, D. (2002). Competency and individual performance: Modelling the world of work. In I. T. Robertson, M. Callinan, & D. Bartram (Eds.), *Organizational effectiveness: The role of psychology* (pp. 227-255). Chichester, UK: John Wiely & Sons.

Latham, G. P., & Whyte, G. (1994). The futility of utility analysis. *Personnel psychology*, *47*(1), 31-46. doi:https://doi.org/10.1111/j.1744-6570.1994.tb02408.x

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological methods*, *4*(1), 84-99. doi:https://doi.org/10.1037//1082-989x.4.1.84

Mansfield, R. S. (1996). Building competency models: Approaches for HR professionals. *Human Resource Management*, *35*(1), 7. doi:https://doi.org/10.1002/(sici)1099-050x(199621)35:1<7::aid-hrm1>3.0.co;2-2

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. Retrieved from https://bigdatawg.nist.gov/pdf/MGI_big_data_full_report.pdf

Marcus, B., & Schuler, H. (2004). Antecedents of counterproductive behavior at work: A general perspective. *Journal of applied psychology*, *89*(4), 647-660. doi:https://doi.org/10.1037/0021-9010.89.4.647

Marcus, B., Taylor, O. A., Hastings, S. E., Sturm, A., & Weigelt, O. (2016). The structure of counterproductive work behavior: A review, a structural meta-analysis, and a primary study. *Journal of Management*, *42*(1), 203-233. doi:https://doi.org/10.1177/0149206313503019

Markus, L., Thomas, H., & Allpress, K. (2005). Confounded by competencies? An evaluation of the evolution and use of competency models. *New Zealand Journal of Psychology*, *34*(2), 117-126.

Marler, J. H., & Boudreau, J. W. (2016). An evidence-based review of HR Analytics. *The International Journal of Human Resource Management*, *28*(1), 1-24. doi:10.1080/09585192.2016.1244699

Mayo, A. (2001). *Human value of the enterprise*. Londen, England: Nicholas Brealey Publishing.

McClelland, D. C. (1973). Testing for competence rather than for intelligence. *American Psychologist*, *28*(1), 1-14. doi:https://doi.org/10.1037/h0034092

Nicolaus, H., Jacques, B., Michael, C., James, M., Tamim, S., Bill, W., & Guru, S. (2016). *The age of analytics: Competing in a data-driven world*. Retrieved from http://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world

Ones, D. S., & Viswesvaran, C. (2011). Individual differences at work. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell handbook of individual differences* (pp. 379-407). Oxford, UK: Wiley-Blackwell.

Paauwe, J., & Boselie, P. (2005). HRM and performance: What next? *Human Resource Management Journal*, *15*(4), 68-83. doi:https://doi.org/10.1111/j.1748-8583.2005.tb00296.x

Pallant, J. (2005). *SPSS survival manual: A step by step guide to data analysis using SPSS for windows version 12*. Syndey, Australia: Allen & Unwin.

Peccei, R., van de Voorde, F., & Van Veldhoven, M. (2013). HRM, well-being and performance: A theoretical and empirical review. In J. Paauwe, D. Guest, & P. Wright (Eds.), *HRM & performance: Achievements & challenges* (pp. 15-45). Chichester UK: Wiley.

Power, D. J. (2014). Using 'Big Data' for analytics and decision support. *Journal of Decision Systems*, *23*(2), 222-228. doi:http://dx.doi.org/10.1080/12460125.2014.888848

Press, G. (2013, 19 August). Data science: What's the half-life of a buzzword? *Forbes*. Retrieved from https://www.forbes.com/sites/gilpress/2013/08/19/data-science-whats-the-half-life-of-a-buzzword/#2758077e7bfd

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta psychologica, 104*(1), 1-15. doi:https://doi.org/10.1016/s0001-6918(99)00050-5

Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of applied psychology, 85*(4), 612-624. doi:10.1037//0021-9010.85.4.612

Rich, B. L., Lepine, J. A., & Crawford, E. R. (2010). Job engagement: Antecedents and effects on job performance. *Academy of management journal, 53*(3), 617-635. doi:https://doi.org/10.5465/amj.2010.51468988

Richard, P. J., Devinney, T. M., Yip, G. S., & Johnson, G. (2009). Measuring organizational performance: Towards methodological best practice. *Journal of Management, 35*(3), 718-804. doi:https://doi.org/10.1177/0149206308330560

Russom, P. (2011). *Big data analytics*. Retrieved from https://tdwi.org/research/2011/09/best-practices-report-q4-big-data-analytics.aspx?tc=page0&tc=assetpg

Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual review of psychology, 59*(1), 419-450. doi:https://doi.org/10.1146/annurev.psych.59.103006.093716

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human performance, 15*(1-2), 187-210. doi:https://doi.org/10.1080/08959285.2002.9668091

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin, 124*(2), 262-274. doi:http://dx.doi.org/10.1037/0033-2909.124.2.262

Services, B. B. R. (2011). *The current state of business analytics: Where do we go from here*. Retrieved from http://www.sas.com/resources/asset/busanalyticsstudy_wp_08232011.pdf

Shavelson, R. J. (2010). On the measurement of competency. *Empirical research in vocational education and training, 2*(1), 41-63.

Shippmann, J. S., Ash, R. A., Batjtsta, M., Carr, L., Eyde, L. D., Hesketh, B., . . . Sanchez, J. I. (2000). The practice of competency modeling. *Personnel psychology, 53*(3), 703-740. doi:https://doi.org/10.1111/j.1744-6570.2000.tb00220.x

Smith, M., & Smith, P. (2005). *Testing people at work: Competencies in psychometric testing*. Malden, MA: BPS Blackwell.

Sonnentag, S., & Frese, M. (2002). Performance concepts and performance theory. In S. Sonnentag (Ed.), *Psychological Management of Individual Performance* (Vol. 23, pp. 3-25). Chichester, UK: John Wiley & Sons.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology, 15*(2), 201-292. doi:https://doi.org/10.2307/1412107

Spector, P. E. (2012). *Industrial and organizational psychology: Research and practice* (6 ed.): John Wiley And Sons.

Spencer, L. M., & Spencer, S. M. (2008). *Competence at work: Models for superior performance*: John Wiley & Sons.

Steen, A., & Welch, D. (2011). Are accounting metrics applicable to human resources? The case of return on investment in valuing international assignments. *Australasian Accounting Business & Finance Journal, 5*(3), 57-72.

Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of personality assessment, 80*(3), 217-222. doi:https://doi.org/10.1207/s15327752jpa8003_01

Taylor, D. (2016, September 21). Battle of the data science venn diagrams. Retrieved from http://www.prooffreader.com/2016/09/battle-of-data-science-venn-diagrams.html

Thorndike, E. L., Bregman, E. O., Cobb, M. V., & Woodyard, E. (1926). *The measurement of intelligence*. New York: Columbia University.

Ulrich, D. (1997). *Human resource champions: The next agenda for adding value and delivering results*. Boston, MA: Harvard Business Review Press.

Ulrich, D., & Dulebohn, J. H. (2015). Are we there yet? What's next for HR? *Human Resource Management Review, 25*(2), 188-204. doi:http://dx.doi.org/10.1016/j.hrmr.2015.01.004

van der Aalst, W. M. (2014). Data scientist: The engineer of the future. In K. Mertins, F. Bénaben, R. Poler, & J.-P. Bourrières (Eds.), *Enterprise Interoperability VI* (pp. 13-26). Albi, France: Springer.

Viswesvaran, C. (2001). Assessment of individual job performance: A review of the past century and a look ahead. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 1, pp. 110-126). London, England: SAGE.

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment*, *8*(4), 216-226. doi:https://doi.org/10.1111/1468-2389.00151

Willems, K. (2015, November 10). The data science industry: Who does what (infographic). *Learning Data Science*.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology*, *9*(2), 79-94. doi:https://doi.org/10.20982/tqmp.09.2.p079

Appendices

Appendix 1. Original Competency Framework from Organization X

Competenties	Items
	Data & Technology
Programmeren	Is in staat om de juiste resultaten te programmeren
	Is in staat om op efficiënte en effectieve wijze te programmeren (technisch slim)
Basic programmeertalen (aantal en welke)	MS Excel
	Business Objects
	MS SQL Server
	Microsoft SQL Management Studio
	Microsoft SQL Analysis software
	Oracle SQL Developer
	Teradata Studio Express
	Teradata SQL Assistant
	MS AccesManager
	SAS
Advanced programmeertalen (aantal en welke)	Large scale database systems
	NoSQL
	HadoopMapReduce
	Hadoop data querying
	Spark
	Cognos
	Knime
	Matlab
	Python
	R
	RapidMiner
	Scala
	Tableau
	Epiphany Campaign Management
	Customer Interaction Manager
Data koppelingen	Is in staat om informatie uit verschillende datasets aan elkaar te koppelen
	Is in staat om datasets tot een hoger niveau te aggregeren/groeperen
Beheersing van data ontsluiting, kwaliteit en omgeving	is in staat de businessvraag te vertalen naar functionele specificaties
	Is in staat om één data en analytics omgeving te realiseren waarin klantdata beschikbaar is voor analyses
	Is in staat om nieuwe databronnen te ontsluiten en in te zetten
	Is in staat om problemen met de datakwaliteit op te sporen en op te lossen
Visie en ideeën over data ontsluiting, kwaliteit en omgeving	Heeft kennis van database-technologie en software en kan participeren in management discussies over hoe een data omgeving gemanaged en gestructureerd moet zijn
	Heeft een visie over hoe een goede analyseomgeving opgebouwd moet zijn

	Visie aanwezig over hoe data dusdanig gestructureerd kan worden dat variabelen (zoals gebruik, kosten, omzet) op klantniveau gevolgd kunnen worden over de tijd
	Analyse methoden en technieken
Analytisch vermogen	Is in staat om essentie uit veelheid van informatie te destilleren Is in staat om verbanden te leggen en te zien
Analyse vaardigheden	Is in staat om business vragen te vertalen naar de juiste analyse-aanpak Is in staat om nieuwe analyse-aanpakken te introduceren en in te zetten Is in staat om met behulp van relevante (statistische) analysemethoden en technieken tot het juiste resultaat te komen Is in staat om tot nieuwe en verrassende analyses en inzichten te komen Is in staat resultaten van (statistische) analysemethoden en technieken juist te interpreteren
Basic Analytics - exploratief (aantal en welke)	Draaitabel Correspondentieanalyse (kruistabel) Profielanalyse Like-4-Like analyse Cohortanalyse Clusteranalyse Textmining technieken Clustering technieken
Basic Analytics - statistiek (aantal en welke)	Factoranalyse Lineaire regressie Logistische regressie Decision tree/Chaid analyse Significantie toetsen Samenhang toetsen Regressiemodellen optimalisatie Wiskundige/statistische technieken programmeren in Python/R/Matlab Tijdseffecten bepalen met gebruik van timeseries Survival analyses
Advanced Analytics (aantal en welke)	Associated Rule Ensemble Bayesian Neural Networks Regularization Instance based Deep learning
	Impact en adviesvaardigheden
Vraagstelling & hypothese formuleren	Is in staat om bij een (analyse) vraag te achterhalen wat de achterliggende doelstelling en/of behoeften zijn Is in staat om hypotheses te formuleren
Leidende rol op het gebied van data & analytics	Is in staat om proactief kansen te signaleren en potentieel te kwantificeren
	Is in staat om analyse-inzichten te vertalen naar een relevante boodschap

Gestructureerd presenteren	Is in staat om de boodschap te vertalen naar een gestructureerde verhaallijn (bijvoorbeeld met behulp van de Pyramid Principle)
	Is in staat om de analyseresultaten te visualiseren op een manier die de verhaallijn versterkt
Communicatieve vaardigheden	Is in staat om vragen te stellen en door te vragen om te achterhalen wat de ander belangrijk vindt/verwacht
	Is in staat om actief te luisteren en weer te geven wat de ander inbrengt
	Is in staat om helder en duidelijk een verhaal te communiceren
	Heeft een open, uitnodigende non-verbale houding
Persoonlijke impact	Is in staat om de output te vertalen in voor de business relevante impact
	Presenteert zelfverzekerd en overtuigend
	Is in staat om de analyse/het project te 'verkopen' met steekhoudende argumenten
	Maakt een krachtige, professionele indruk op anderen
Plannen & Organiseren	In staat om activiteiten op gestructureerde wijze te plannen en te organiseren.
	Is in staat om urgent en belangrijk te onderscheiden
	In staat om randvoorwaarden te creëren (datatoegang, capaciteit, middelen, etc).
Samenwerken	In staat om te zorgen voor heldere afstemming en overleg met collega's/klanten
	Is in staat om met anderen op constructieve wijze gemeenschappelijke doelen te bereiken
	Is in staat om op constructieve wijze feedback te geven
	Is in staat om anderen te motiveren/stimuleren
Resultaat & doelgerichtheid	Is in staat om binnen besproken deadlines te zorgen voor zichtbaar resultaat
	Is in staat om mensen aan te sturen zodat volgens tijd, kwaliteit en kosten impactvolle output wordt geleverd
	Is gedreven om zijn concrete doelen en resultaten te bereiken
Stakeholder/Verwachtingsmanagement (hier of onder impact- en adviesvaardigheden)	Is in staat om zich in te leven in en te reageren op de behoeften van een (interne of externe) klant
	Is in staat een goede relatie op te bouwen en te onderhouden met stakeholders
	Is in staat om de verwachtingen (van stakeholders) te managen
	Is in staat om een gedegen implementatieplan op te stellen om de analyse-output te implementeren en te borgen in de organisatie
	Is in staat om een gedegen implementatieplan op te stellen om de analyse-output te implementeren en te borgen in de organisatie
Business Domain Expertise	
Ontwikkelen, meten en verbeteren initiatieven	Is in staat om het effect van een actie/initiatief te meten en monitoren
	Is in staat om bij het analyseren van uitkomsten van een actie/initiatief op zoek te gaan naar verbeterkansen middels het uitvoeren van verdiepende analyses
	Is in staat tot het formuleren van concrete verbeterideeën over acties/initiatieven op basis van (analyse-, data-, of campagne-) resultaten

	Is bekend met de visie en strategie van de organisatie en weet waar de speerpunten voor de komende periode liggen
Kennis van de business	is bekend met de bedrijfstak waarin de organisatie opereert en weet hoe de organisatie daarin presteert ten opzichte van de concurrentie
	Is bekend met de proposities en producten van de organisatie
	Is in staat om (verwachte) klantwaarde in te schatten
Klantwaarde	Is in staat om een ROI berekening te maken
ROI	Is in staat om een forecasting model te maken
Forecasting	Is in staat om inzichten te genereren vanuit online data
Online/digital	Is in staat om pricing modellen te bouwen
Pricing	In staat om customer journey inzichtelijk te maken
Customer journey	In staat om vanuit inzichten pro-actief met verbeterinitiatieven te komen die breder in de business inzetbaar zijn
Impact creeren op business	Bekend met concrete doelstellingen van business partner en in staat hierop aan te sluiten
doelstellingen	In staat om korte termijn impact en (langere termijn) innovatie in balans te houden in werkzaamheden

Appendix 2. Individual Work Performance Questionnaire (Koopmans et al. 2014)

The following questions relate to how you carried out your work during the past 3 months. If you are uncertain about how to answer a particular question, please give the best possible answer. This part will take about 3-5 minutes to complete. The questionnaire is completely anonymous: your answers will not be seen by your supervisor(s) or colleagues.

Seldom Sometimes Regularly Often Always

In the past 3 months… TP

1. I was able to plan my work so that I finished it on time.
2. I kept in mind the work result I needed to achieve.
3. I was able to set priorities.
4. I was able to carry out my work efficiently.
5. I managed my time well.

In the past 3 months… OCB

1. On my own initiative, I started new tasks when my old tasks were completed.
2. I took on challenging tasks when they were available.
3. I worked on keeping my job-related knowledge up-to-date.
4. I worked on keeping my work skills up-to-date.
5. I came up with creative solutions for new problems.
6. I took on extra responsibilities
7. I continually sought new challenges in my work
8. I actively participated in meetings and/or consultations

Never Seldom Sometimes Regularly Often

In the past 3 months… CWB

1. I complained about minor work-related issues at work.
2. I made problems at work bigger than they were.
3. I focused on the negative aspects of situation at work instead of the positive aspects.
4. I talked to colleagues about the negative aspects of my work.
5. I talked to people outside the organization about the negative aspects of my work.

Appendix 3. Feedback on Competency Framework Study 1

Onduidelijke items

- "(20) Ik vind het een aparte lijst. Een deel zijn machine learning technieken, zoals deep learning, maar bijv. regularization is een wiskundige 'truc' in o.a. deep learning om overfitting tegen te gaan."
  - Geen concrete feedback voor verbetering. In overleg besloten om dit niet aan te passen.
- "22: waarop? bij 23 wordt dit gespecificeerd"
  - Item is toegevoegd door trainer van impact & adviesvaardigheden. Is blijkbaar niet helder. Er is een item met visie op data & technology, dit item is meer voor data & analytics.
    - Toevoegen aan item 22: "…op data & analytics". Item 22 en 23 omgedraaid ivm volgorde.
- "20b: Ensemble methodes zijn wel erg breed, hier valt Random Forest bijvoorbeeld ook onder, die misschien wel apart genoemd mag worden door de populariteit binnen MIcompany projecten."
  - Random forest is veel gebruikt, er zijn meerdere methodes. Voor duidelijkheid rf noemen als voorbeeld
    - Toevoegen aan item 20b: "…(bijv. random forest)"
- "Zou even kijken naar de kolomtitels? Binnen bijv. de programmeeromgevingen miste ik een kop 'Geen ervaring' --> Beginner/amateur zijn nagenoeg hetzelfde, dus je kan de 7-puntsschaal prima behouden."
  - Geen ervaring geeft inderdaad duidelijk een 'nul-niveau' aan.
    - Feedback overgenomen en veranderd in de schaal. "Geen ervaring" wordt 1, "beginner" wordt 2, "amateur" is verwijderd.
- "Onder 16 en 17 zijn de termen nieuw en verassend nogal persoonlijk"
  - 'Nieuwe' analyse-aanpakken en inzichten zijn inderdaad verschillend te interpreteren. In item 16 gaat het om innovatieve aanpakken. Item 17 gaat om nieuwe inzichten voor de klant.
    - Veranderen in item 16 "Nieuwe" > "innovatieve". Toevoegen aan item 17: "…voor de klant"
- "Vraag 53 t/m 59 zijn vrij algemeen geformuleerd. Gaat dit dan over kennis over de business van MIcompany of van de klant? Van sommige klanten ken ik ze goed, maar van sommige veel minder. Daarom is deze vraag vanuit consultancy lastig te interpreteren en in te vullen."
  - Is inderdaad moeilijk voor externe analisten. Studie 2 gaat enkel om interne analisten.

Ontbrekende competenties

- "Version control is essentieel in een data-driven analist."
  - Goede suggestie, maar nog onduidelijk. Gemaild om opheldering te vragen:
  - "Version Control valt onder data en technology. Het gaat om het gebruik van software (Git bijvoorbeeld) die het werken aan code door verschillende mensen tegelijk makkelijk maakt, en ervoor zorgt dat veranderingen teruggedraaid kunnen worden. Goed gebruik maken hiervan is essentieel voor

het ontwikkelen van database suites en softwarepakketten, en aan te raden voor het ontwikkelen van analysescripts in bijvoorbeeld R.”

- **Item toevoegen over version control**

- “Er missen wat machine learning technieken, maar ik weet niet hoe relevant je het vindt om die toe te voegen.”
  - Geen concrete feedback.
- “Ervaring met Powerpoint (en Thinkcell)”
  - Hoort bij visualisatie. Het gebruik van het powerpoint is minder relevant dan de visualisatie an sich. Dit kan in meerdere programma’s gedaan worden. Kennis van powerpoint is daarbij minder relevant.
- “Zitten heel veel goede competenties in, volgens mij alle.”
  - Geen concrete feedback.
- “Pricing ontbreekt bij de analytische toepassingen (nur 60 en verder)”
  - Soms heeft een organisatie een aparte pricing afdeling. Dit betekent zeker niet dat dit niet relevant is voor analisten. Besloten om item toe te voegen. Formulering van item voorgelegd aan een program manager.
    - Pricing item toevoegen na item 65. ‘Prijsgevoeligheidsmodel bouwen’

Indeling van competentiedomein en/of items

- “Ik zou de uitsplitsing van advanced analytics als volgt maken: regressie (bv. lineair), classificatie (bv. log regressie, decision trees), dimensie reductie (bv. factor analyse), clustering (bv. k means)”
  - Feedback gaat niet over advanced analytics, maar over een combinatie van exploratief en statistiek basic analytics. Met deze indeling heeft ieder onderdeel maar 1 a 2 items en vallen veel items nergens onder. Besloten om dit niet aan te passen. Bij het reviewen besloten om 18h te verwijderen, dit lijkt sterk op f.
    - 18h verwijderen.
- “De competenties rondom presenteren en intakes staan nu verspreid over verschillende subonderwerpen, die zou ik samenvoegen (bijv 29, 30, 47 onder het achterhalen van de vraag)”
  - Goede feedback. Communicatieve skills en klantfocus kunnen worden samengevoegd onder: ‘klantcommunicatie’. Alleen item 52 lijkt hier niet onder te passen, deze gaat om resultaten en kan bij resultaat & doelgerichtheid.
    - In overleg communicatieve skills en klantfocus samengevoegd tot klantcommunicatie en item 52 bij r&d gezet.

Andere suggesties

- “Vraag niet naar itemnummers, je zult weinig personen vinden die dit gaan onthouden. Voor de rest interessant onderwerp, zet mij ook aan het denken. Succes!”
  - Gaat over de feedbackvragen. Er werd verwezen naar de vragenlijst die in zijn geheel onder de vragen afgebeeld stond. Heeft deze persoon waarschijnlijk gemist.
- “Ik zou nog een stukje toelichting geven bij elk van de 4 data science domains. Ook vind ik de vragen over "Business domain expertise" lastig. Een sterke expertise in bijv. retail betekent niet automatisch dat je ook expertise in andere sectoren hebt, dus hoe vul je de vragen dan in?”

  - Is inderdaad moeilijk voor externe analisten. Zelfde vraag als voorheen.
- “De vraag wat is belangrijk aan het einde is lastig om te beantwoorden, omdat er te veel is om uit te kiezen. Je kan misschien beter per domein vragen welke 2 het belangrijkste zijn?”
  - Gaat over de feedbackvragen. Het is juist interessant om te zien of bepaalde domeinen meer of minder relevant lijken te zijn, daarom vrije keuze.
- “Ik zou het prettig vinden om een structuurboom te zien van het competentiemodel. Er zijn >20 onderdelen en de kapstok bestaat maar uit vier domeinen.”
  - Geen concrete feedback. Stond onderaan de vragen, hier werd naar gerefereerd.
- “Aangeven wat het percentage afgerond is.”
  - Impliceert dat je een competentie 100% kunt hebben. Hierbij zal interpretatie een groot probleem zijn. Niet aanpassen.
- “Ik vraag mij af of beginner tot expert bij draagt, het vult makkelijker in op likert schaal en wellicht minstens zo goed?”
  - In verband met het anchoring probleem gekozen om deze niveaus in de schalen te laten staan. Niet aanpassen.
- “Sorry dat ik de open vragen hier vlak voor niet heb ingevuld, maar ik vind het te moeilijk om uit 70 competenties er 8 te kiezen. Dat zou behoorlijk random worden wat mij betreft. Volgens mij is het voor een analist belangrijk om nadat je een basisniveau op alle vlakken hebt, te bepalen wat je specialisme wordt; statistische methoden, data technology, etc. en daar vervolgens expert in te worden. Hopelijk helpt dit antwoord je.”
  - Belangrijke feedback. Een goede analist hoeft niet per se alle programmeertalen te bezitten, maar zal zich specialiseren. Hoe kun je dan het niveau van de programming competenties beoordelen? Overal een beetje van weten is mogelijk minder waardevol dan op een gebied bedreven/expert zijn. Hoe bepaal je dan de score hiervan?
    - Bij scoring van 9, 10, 18, 19, 20 alleen de scores vanaf niveau 4 (geavanceerd) meenemen. 1,2,3 > 0; 4,5>1; 6,7>2.

Alle veranderingen zijn ook toegepast op de Engelse vragenlijst.

Feedback English questionnaire

- Change rating scale: Lower advanced > intermediate
- Item 3. Merging > collecting
- Item 6. one > a
- Item 7. > ask translation for ontsluiten. Senior analyst suggested ‘extracting’
- Add another item in general D&T skills about visualization
  - ‘Visualizing results in a simple and concise manner’
- Restructure the basic/advanced parts
  - New categories: Databases, analytical modeling, programming languages, big data, general data tool, business intelligence
  - Leave out: Tableau (visualization item is included in general skills), 10 ‘g, k, n, o’ (these are small tools), 10a (this is very general and doesn’t fit here)
  - Add analytical modeling programs, SPSS and Stata.
  - Add programming language Java

  - Ask for other programs that are not in the list, because this list is not exhaustive.
- Item 11. Reformulate this item: Obtaining the meaning from large data sets
- Item 12. Insert ‘relevant’: Making relevant connections
- Item 13. approach > method
- Item 14. Applying new analytical methods
- Item 17. generating > producing. And/or unexpected
- Item 18.a leave out
- Leave out the words ‘basic’ and ‘advanced’. Combine 18 & 19 in one table, this will become ‘statistics’.
- Leave out 19 e, h
- Item 19.i time series analyses.
- Item 20b,c Ensemble & Bayesian could be replaced to the ‘statistics’ table.
- Item 21 put proactively at beginning
- Item 28 to realize the desired goal
- Item 29 into > in to
- Item 36 and > in
- Item 37 Premise > Prioritize
- Item 41 right > appropriate
- Item 43 implemented > obtained
- Item 44 in depth > in-depth
- Item 71 in depth > in-depth

Appendix 4. Competency Questionnaire Study 2

Demographics

-Gender: Male Female
-Age: <26, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, >60
-Education: What is the highest degree or level of school you have completed?
  High school/vocational training
  Associate degree
  Bachelor’s degree at university
  Master’s degree at university
  PhD or other advanced degree beyond a Master's degree
-Current function title:

Competency questionnaires

Below you find questions regarding your current level of several competencies. These competencies are grouped in four categories. The questionnaire is anonymous.
Are you unsure which answer you should fill in? Please indicate the best possible answer.

Scale

1. **No experience** – No experience/ no or little knowledge
2. **Beginner** – Familiar with basics/ difficulties with applications/ requires much guidance
3. **Intermediate** – Experience with applications of basics/ knowledge about advanced applications/ requires guidance
4. **Advanced** – Experience with advanced applications/ requires guidance occasionally
5. **Competent** – Independently executing advanced applications
6. **Proficient** – Applying to new (complex) situations/ helping others
7. **Expert** – Full competence/ vision about developments/ role model for others

Data and technology competencies (1/4)
Please indicate to what extent you possess the following competencies
General data and technology competencies

1. Programming the right results
2. Programming in an efficient and effective way (technical smart)
3. Collecting information from diverse datasets
4. Aggregating/grouping datasets to a higher level
5. Translating the business question to functional specifications
6. Creating a data and analytics environment in which customer data is available for analyses
7. Extracting and using new data sources
8. Detecting and solving problems with data quality
9. Visualizing results in a simple and concise manner in a dashboard*

10. Databases
    a) SQL
    b) MS AccessManager
    c) NoSQL database (any sort)
11. Analytical modeling
    a. SAS
    b. SPSS
    c. Stata
12. Programming languages
    a. R
    b. Python
    c. Matlab
    d. Scala
    e. Java
13. Big data
    a. Hadoop
    b. Spark
14. General data tools
    a. MS Excel
    b. Teradata studio express
15. Business intelligence
    a. Cognos
    b. Business Objects
16. Are there any competencies missing in the lists above that you find essential for a good data analyst? If yes, which one(s)?

Analysis methods and techniques (2/4)
Please indicate to what extent you possess the following competencies
General analysis methods and techniques competencies

17. Obtaining the meaning from large data sets
18. Making relevant connections
19. Translating the business question to the right analytical method
20. Applying analytical methods to get the right results

21. Interpreting results correctly
22. Introducing and using innovative analysis approaches
23. Producing insights that are new and/or unexpected to the customer
24. Statistics
    a. Correspondence analysis (cross Table)
    b. Profile analysis
    c. Clustering analysis
    d. Factor analysis
    e. Linear regression
    f. Logistic regression
    g. Decision tree/Chaid analysis
    h. Coherence test
    i. Regression model optimization
    j. Time series analysis
    k. Survival analysis
25. Machine learning
    a. Associated Rule
    b. Ensemble (e.g. random forest)
    c. Bayesian
    d. Neural Networks
    e. Regularization
    f. Instance based
    g. Deep learning
26. Other analyses
    a. Like-4-Like analysis
    b. Cohort analysis
    c. Text mining techniques

27. Are there any competencies missing in the lists above that you find essential for a good data analyst? If yes, which one(s)?

Impact and advisory skills (3/4)
Please indicate to what extent you possess the following competencies
Leading role in the field of data & analytics

28. Proactively signaling opportunities and quantifying potential
29. Defining and presenting personal vision on data & technology*
30. Defining and presenting personal vision on data & analytics*
31. Steering people to deliver impactful output

Structured presentation

32. Translating insights from analyses to a relevant message
33. Translating the message to a structured storyline
34. Visualizing results from analyses in a way that enhances the storyline
35. Presenting the storyline in a convincing way to realize the desired goal

Personal impact

36. Translating the output in to impact that is relevant for the business
37. Presenting in a confident and convincing way
38. ‘Selling’ the analysis/project with strong arguments

39. Making a strong, professional impression on others

Planning and organizing

40. Planning and organizing activities in a structured way
41. Differentiating between major and minor issues
42. Creating conditions for performing analyses (data access, capacity, resources, etc.)

Team work

43. Creating clear agreements in consultation with colleagues/customers
44. Prioritize shared goals above personal goals
45. Providing constructive feedback
46. Motivating/stimulating others

Results & goal orientation

47. Providing visible results within agreed deadlines
48. Deliver impactful output with the appropriate standards of time, quality, and costs
49. Driven to reach concrete goals and results
50. Composing a thorough implementation plan so that results can be obtained

Customer communication

51. Asking questions and going in-depth to understand the priorities and expectations
52. Active listening, summarizing and picturing the input from others
53. Clearly communicating a story
54. Having an open, inviting, non-verbal appearance
55. Understanding the underlying goals and/or needs with an (analytical) question
56. Formulating hypotheses concisely
57. Empathizing with customers and responding to their needs
58. Building and maintaining good relations with stakeholders
59. Managing (stakeholders') expectations

Business domain expertise (4/4)

Please indicate to what extent you possess the following competencies

Knowledge of the business

60. Knowledge of the business functions/units*
61. Knowledge of the business strategy and vision
62. Knowledge of propositions and products
63. Knowledge of current problems and opportunities
64. Knowledge of the industry
65. Knowledge of the competitive position
66. Knowledge of important stakeholders

Analytical applications

67. Providing suitable analytical solutions
68. Estimating the (expected) customer value
69. Making ROI calculations
70. Making a forecasting model
71. Generating insights from online data
72. Making customer journey insightful

73. Making a pricing sensitivity model*

Creating impact on business goals

74. Thinking of broad applicable initiatives
75. Providing suitable solutions to the business question
76. Generating both short term impact and (long term) innovation

Implementing initiatives

77. Measuring and monitoring the effect of an action/initiative
78. Detecting further improvement opportunities through in-depth analyses
79. Formulating concrete improvements based on (analysis, data, or campaign) results

Appendix 5. New Competency Questionnaire with Short Scales

Dimensions		α	Item#	Competencies
DT	Programming	.920	DT1	Programming the right results
			DT3	Collecting information from diverse datasets
			DT4	Aggregating/grouping datasets to a higher level
DT	Architecture & quality	.852	DT6	Creating a data and analytics environment in which customer data is available for analyses
			DT7	Extracting and using new data sources
			DT8	Detecting and solving problems with data quality
DT	Program index		DT10	Databases
			DT11	Analytical Modeling
			DT12	Programming languages
			DT13	Big data
			DT14	General data tools
			DT15	Business intelligence
			New	Visualization
AMT	AMTgeneral	.957	AMT17	Obtaining the meaning from large data sets
			AMT18	Making relevant connections
			AMT19	Translating the business question to the right analytical method
			NEW	Playing devil's advocate/ question analyses and results
AMT	Basic analytics	.961	AMT24-efi	Regression techniques (e.g. linear, logistic, model optimization)
			AMT24-ab26b	Explorative techniques (e.g. profile-, cohortanalysis)
			AMT24c	Clustering analysis
			AMT24g	Decision tree/Chaid analysis
AMT	Advanced analytics	.870	AMT25e	Regularization
			AMT25f	Instance based
			AMT25a	Associated Rule
IA	Planning & interaction	.946	IA57	Empathizing with customers and responding to their needs
			IA42	Creating conditions for performing analyses
			IA43	Creating clear agreements in consultation with colleagues/customers
			IA56	Formulating hypotheses concisely
			IA55	Understanding the underlying goals and/or needs with an (analytical) question
			IA44	Prioritize shared goals above personal goals
			IA48	Deliver impactful output with the appropriate standards of time, quality, and costs
IA	Presenting results	.965	IA35	Presenting the storyline in a convincing way to realize the desired goal
			IA33	Translating the message to a structured storyline
			IA34	Visualizing results from analyses in a way that enhances the storyline
			IA32	Translating insights from analyses to a relevant message
			IA37	Presenting in a confident and convincing way

BDE	Business knowledge	.937	BDE66	Knowledge of important stakeholders
			BDE60	Knowledge of the business functions/units
			BDE61	Knowledge of the business strategy and vision
BDE	Analytical applications	.905	BDE73	Making a pricing sensitivity model
			BDE70	Making a forecasting model
			BDE69	Making ROI calculations
			BDE68	Estimating the (expected) customer value
			BDE67	Providing suitable analytical solutions
			BDE72	Making customer journey insightful