Purchasing Assessment Tool – Assessment and benchmarking of strategic purchaser skill set

Master Thesis

by

Harald Österling
for
The department of Industrial Management and Logistics
advised by
Louise Bildsten, Faculty of Engineering, Lund University and Björn Munkhammar, Procurement Power
and examined by
Jan Olhager, Faculty of Engineering, Lund University

Lund University
Lund, Sweden
June 2022

Abstract

According to studies, a strategically aligned purchasing strategy can increase both purchasing and business performance. The foundation for executing a strategically aligned purchasing strategy is the fact that strategic purchasing professionals inherit the relevant competencies and that organizations have tools to track purchasing professionals' performance. However, these performance- and competency measurements are usually customized to a specific purchasing function, and therefore there are few standardized and objective ways for assessing purchasing competencies.

This master thesis aims to create a standardized Multiple Choice Question competence assessment tool for measuring relevant competence areas. As a first part, seven relevant competence areas are identified through literature study, archival study of educational content and interviews. Secondly, a competence assessment tool is created and validated on 143 purchasing professionals from 13 Swedish companies. The seven chosen competency areas are Sourcing Strategy, Negotiation, Contracting & Contract Management, Cost & Value Mangement, Risk Management, Sustainability, and Financing. The findings from the empirical study suggest that the competency areas are relevant for purchasing professionals, and that the purchasing assessment tool is a valid and reliable form of measuring competencies.

This thesis focuses solely on strategic purchasing professional competencies, and does not consider any company-specific organizational factors. It concludes with a discussion on future research.

Keywords: Purchasing and supply management, Procurement, Competencies, Purchasing performance management (PPM), Strategic purchasing competencies, Competency assessment

"Supply can only be strategic if the people working in that area possess the requisite skills and competencies to operate in that way."
- Cousins et al. (2008)

Acknowledgements

I would like to give my warmest thanks to all colleagues at Procurement Power for the expertise required to conduct this master thesis. During my months at their office I had the opportunity to discuss and iterate ideas, which was imperative for finishing the Competence Assessment Tool. I would also like to thank Louise Bildsten for the guidance and willingness to help with hour-long meetings and interviews. I would like to warmly thank my family and my partner who have become involved in the thesis for discussions and such, whether they liked it or not. Finally, I would like to thank all the participating strategic purchasers, purchasing managers, and course lecturers for taking the time and enabling a more extensive empirical study than I could have hoped for.

Harald Österling, 1 June 2022

Definitions

MCQ: Multiple choice question

MCQ test case: A question in the MCQ test

PPMS: Purchasing Performance Measurement System

Stem: The question or statement that should give the stimulus to an answer.

TCO: Total Cost of Ownership

Key: Correct answer

Distractor: Incorrect answer

Competencies: The strategic and analytic skills and abilities to make effective and efficient judgments based on data in a purchasing setting. Synthesis of definitions by Bals et al. (2019) and Cousins et al. (2008).

Contents

List of Figures

List of Tables

1

Introduction

1.1 Background

Most companies today spend 50 percent or more of their annual budget on purchasing (Van Weele, 2009). However, purchasing has until recently primarily been seen as an operational role with little strategic importance (Van Weele, 2009; Ellram and Carr, 1994). In the recent decades, the strategic purchasing function and the strategic purchasing professionals' role of managing the external connections upstream within the supply chain has become more important for competitive advantage (Tassabehji and Moorhouse, 2008; Baier et al., 2008), Carr and Smeltzer (1999); Cousins et al. (2008), and several empirical studies has shown that high-performing strategic purchasing enables long-term profit advantages and a considerably faster recovery from financial crises and economic downturns (Bayazit et al., 2015; Carr and Smeltzer, 1999; Luzzini and Ronchi, 2016).

Making sure that employees inherit the relevant individual skills is suggested to be relevant for sustaining performance and strategic advantage over time (Bayazit et al., 2015; Bals et al., 2019) and senior executives are placing a greater emphasis on developing purchasing competencies (Bayazit et al., 2015). Eilström and Kock (2008) emphasizes three useful applications of assessment for professional development. The first is career planning such as promotion and career mobility. The second is employee training, and the third is job task development to expose employees to new competency requirements.

The purchasing strategy of firms is implemented in purchasing departments through practices and activities that should be aligned with the purchasing strategy (Pohl and Förstl, 2011; Cousins et al., 2008). The success of implementing the purchasing strategy is called the purchasing performance, and this is measured through purchasing performance measurement systems (PPMS) companies in order to influence behavior and communicate the role of the purchasing function (Pohl and Förstl, 2011; Cousins et al., 2008).

1.2 Research problem

Although PPMS are output related, they are often used as diagnoses for competency development (Pohl and Förstl, 2011; Aulia and Isvara, 2021). Due to different competitive priorities of different companies, the priorities are shown to vary quite a lot for performance measurement, making them impractical to use for benchmarking purposes (Caniato et al., 2014). Aulia and Isvara (2021) and Bayazit et al. (2015) suggest that it is relevant to be able to assess certain important competencies among strategic purchasers in order to increase the purchasing performance of the firm.

This thesis aims to contribute to enabling a higher purchasing performance by creating an objective and data driven assessment by creating a tool for benchmarking competencies between purchasing professionals across industries.

1.3 Purpose and Research Questions

1.3.1 Purpose

The purpose is to create and validate a tool that measures the extent to which strategic purchasers have some competencies relevant to high performing purchasing and supply chain organizations.

1.3.2 Research Questions

- How can a suitable number of competency areas that are relevant for the performance of strategic purchasing professionals be identified?
- How can a tool be created for an objective and structured assessment of these competency areas? Henceforth called a purchasing assessment tool.
- How can the validity and reliability of the purchasing assessment tool be extensively assessed?

1.4 Focus and delimitation

The master thesis mainly focuses on the design, execution, and validation of the purchasing assessment tool. The thesis analyzes the relevant characteristics of purchasing professionals in high-performing purchasing organizations and not focus on analyzing the organizational conditions regarding maturity, alignment, technology, etc of the companies. The tool will be delimited to assessing competencies relevant to strategic purchasing professionals.

Purtilo et al. (2015) discusses soft and hard skills with hard skills being specific and teachable skills for completing certain tasks. The Danish Purchasing and

Logistics Forum mentions functional and business skills, with functional skills being about making the right choices based on the information provided (DILF, Accessed 2022). In the strive for finding areas that can both be measured and improved upon, the goal is to focus on hard and functional skills.

1.5 Ethics

Assessment of competencies within ones field of work is a sensitive topic to many participants, and both exposure within the company and towards the author may be discouraging. The same goes for participating companies exposing their overall competence outside of the organisation. The participants of the pilot study therefore remained anonymous towards the author through anonymous codes that the participating companies knew the key to. The companies decided themselves if they wanted the managers to receive all participants' individual scores or if they would only receive a breakdown over the the overall score. All company names also remain anonymous in this study.

1.6 Report Structure

The thesis consists of six chapters:

- 1. Introduction
- 2. Methodology
- 3. Theory
- 4. Empirical Study
- 5. Analysis
- 6. Discussion
- 7. Conclusion

In order to create an assessment tool that is suitable for a certain area of work, the research problems are approached one after the other. Before designing the test altogether, a handful of relevant competency areas will be identified. Once this is done, an assessment of these competency areas can be designed and validated. Since the two parts are distinct from each other, the methodology and theory chapters will be divided up into two parts. Part one will be centered around identifying the relevant competency areas, and part two will be centered around the design and validation of the purchasing assessment tool. The empirical study chapters will bring the two together through since the defined competency areas will be validated mainly through reviewing the contents of the test itself.

2

Methodology

The methodology consists of two distinct parts - determining relevant competency areas for strategic purchasers and designing a test to measure competencies. Due to the different nature of the first and second part of the study, the methodology is divided up into two parts that are both be taken into consideration separately in all layers of the research onion (Saunders et al., 2009). The innermost layer is divided up into two sections in this chapter.

The data collection and analysis of data is greatly affected by choice of for example overall research philosophy, choice of methodology, and research strategy. These aspects should be defined in the planning of a scientific study. Saunders et al. (2009) describes the "Research Onion" seen in figure 2.1 which is a framework for defining a fully covered methodology in a structured manner. This chapter will describe a suitable methodology for the thesis by covering each layer of the research onion.

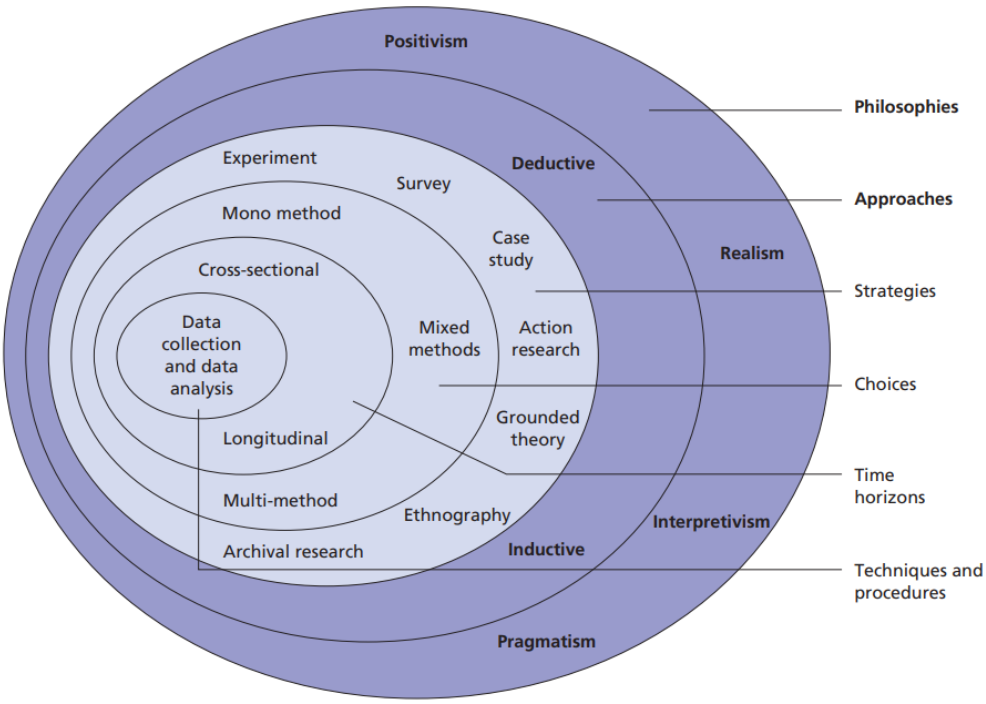


Figure 2.1: Research onion (Saunders et al., 2009).

2.1 Research Philosophy

First off, the researcher should to define what attitude to have towards new knowledge, i.e. research philosophy. The philosophy should be chosen based on what is suitable to the particular research being conducted (Saunders et al., 2009). In management research, Saunders et al. (2009) describes four main philosophies - positivism, interpretivism, realism and pragmatism. Positivism is common in natural sciences since the data gathered is interpreted objectively,leaving little room for own interpretation and assumptions. This philosophy is suitable for quantitative tests, with large amounts of data and clear measurements. On the other side of the spectrum, there is the interpretivism philosophy saying that the gathered data must considered in relation to the context and individual.

The realism approach is a hybrid of positivism and interpretivism. In realism philosophy, direct and critical realism are often differentiated. The critical realist thinks that our senses frequently deceive us while the direct realist does not (Saunders et al., 2009). As a result, critical realism usually examines a situation in two stages. First observing the outcome, and then comprehend the underlying relationships and influences. Finally there is the pragmatism philosophy, stating that there must not be one single approach being chosen (Saunders et al., 2009).

For this study, there are two different research parts. Since the first part of identifying and validating the competency areas has a more qualitative focus, it is suitable to argue that the findings are neither objectively true nor fully subjective. The critical realism approach is therefore suitable for assessing the a hypotheses of relevant competencies, e.g. part 1. For the second part, the validity and reliability analysis of the MCQ test cases consists of a mix of qualitative validation and purely statistical tests for determining the validity and reliability of the test cases. The more qualitative tests such as the content review are also highly structured and standardized to cover all test cases. For the statistical measures, a positivist approach is more suitable. The qualitative reviews should here be taken in a mostly positivist, but somehow critical realist manner, trusting the input of the experts but still thinking critically.

2.2 Research approach

Once the research philosophy is defined, the research approach should be determined. There are three main approaches are the inductive, deductive and abductive approach (Saunders et al., 2009). The inductive approach means making observations and creating theory from there. The deductive approach is on the other hand empirically testing hypotheses that are derived from theory. The abductive approach is a mix between the two, where both inductive and deductive approach is conducted iteratively to both build and test theory (Saunders et al., 2009).

The first part leads to hypothesized competencies that are validated, and the second leads to hypothesized test cases that are validated. The competencies in

the first part are identified through literature review, and the data gathering acts as a validation of these hypotheses. Therefore this part can be seen as deductive. The second part further dives into a the contents of each area, and will later be validated through rigorous statistical testing. However, here the suitable test are reviewed through interviews and discussions in the design phase, leading to valid test cases. Therefore the abductive approach is more suitable.

2.3 Research strategy

The next step is to define the research strategy. This is the first step in setting up the research design, i.e. how the research questions should be answered (Saunders et al., 2009). It is possible to choose several research strategies (Saunders et al., 2009), and in this study this is necessary.

In the first part, the relevant competencies are identified through exploration of what different sources find most relevant. This part identifies both activities and objectives of the strategic purchasers by looking at the existing literature on important competencies. It is validated through analyzing the relevant competencies based on course offerings in strategic purchasing and interviewing industry experts. This takes on a descriptive research strategy, trying to portray what competencies are currently regarded as most relevant (Saunders et al., 2009). Since it is centered around empirical exploration of the concept of a best practice purchasing professional using several sources, a multiple case study is a suitable strategy in this case where the course contents from several universities are analyzed (Saunders et al., 2009).

The second part is of a more explanatory nature, hypothesizing connections between competencies and test cases, and looking into the relationships between purchasing competency and test score through rigorous validity and reliability measures (Saunders et al., 2009). This part of the study takes on a survey based research strategy as it assesses the actions of several different participants getting the same questions in the same controlled environment with highly comparable results (Saunders et al., 2009). Survey based research works well for explanatory and deductive research (Saunders et al., 2009). The answers are quantitative and require statistical analysis. In this case, there is theory readily available of the type of analysis needed in this test (Considine et al., 2005). More information on the specific analysis methods is presented in section 2.7.

2.4 Research Choice

The next step is the research choice. For the first part of finding competency areas, mixed qualitative studies is chosen, which means that several qualitative methods are combined (Saunders et al., 2009). This is suitable since the proposed competency areas are reviewed through cross-case analysis interviews as well as interviews. For the second part, mixed methods is chosen, which means that both qualitative and quantitative data is used (Saunders et al., 2009). This is suitable

since several validity and reliability measures are used, both qualitative and quantitative.

2.5 Time Horizons

The purpose of defining the time horizon is to determine if the research is meant to assess the situation during a certain time/time frame, or if it is meant to test a change over time (Saunders et al., 2009). In this case, the competency areas are both determined and tested in one time frame, instead of the change in relevant areas or competency levels, and therefore the time horizon of this study is cross-sectional (Saunders et al., 2009).

2.6 Techniques and procedures part 1: Determining relevant competency areas for strategic purchasers

The aim of the first part is to define and validate a handful of competencies that are central for strategic purchasers, as well as their contents, with the delimitation of hard or functional competencies taken in mind. The attitude is, of course, that it is not possible to capture all competencies that are crucial for a strategic purchaser and the purpose is therefore to find a handful of relevant competencies. The previous layers have resulted in the possibility of using the Multiple-Case Study Procedure for finding analyzing relevant competency areas for strategic purchasers. Based upon the structure described by Yin (2009), the overall structure for the first part of this study is presented in figure 2.2.

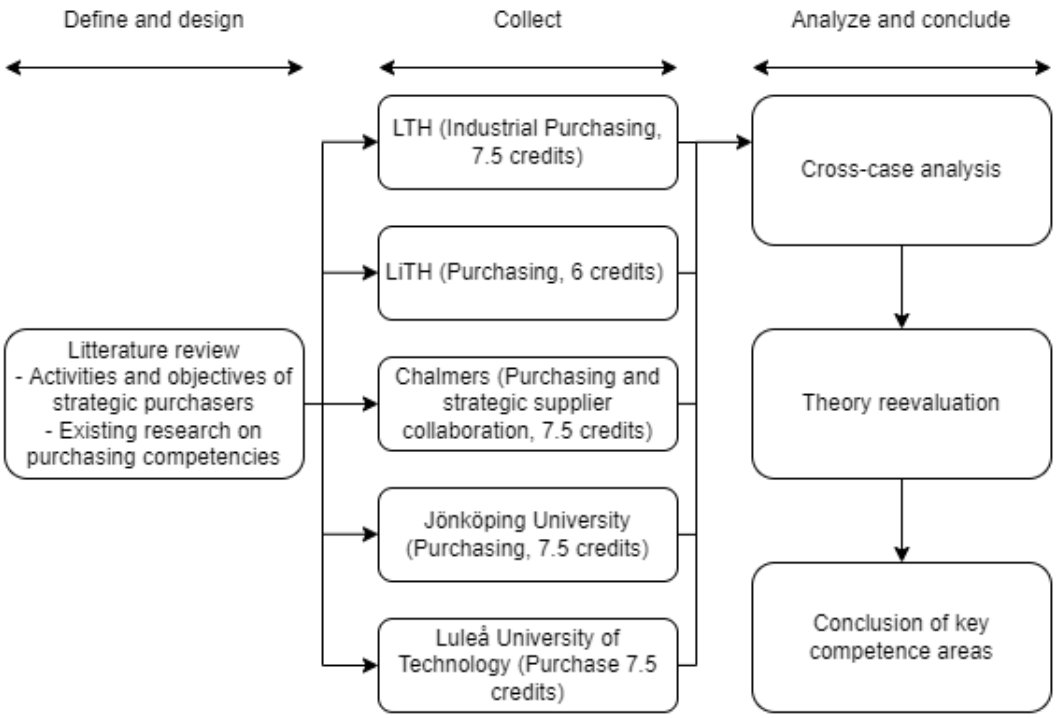


Figure 2.2: Process for finding relevant competency areas.

Firstly, an understanding of what relevant competencies in strategic purchasing professionals are linked to purchasing performance were found through literature review. This lead to a segmentation into a suitable number of seven competency areas that were analyzed through archival study and discussions with experts in the field. The literature review conducted in this section was a semi-systematic review, tracking similar topics and how they have evolved over time (Snyder, 2019). For the specific competency areas presented below, the literature review focused on searching for keywords and synthesizing an area from different articles.

2.6.1 Theory development: Define and design

Firstly, a literature review was conducted to investigate current research on relevant purchasing competencies. The objective was to investigate the strategic purchasing role and objectives as well as current research of competencies and best practices. The literature study focused on tree main areas. Firstly, the strategic purchasing roles and objectives were investigated through descriptive literature. Secondly, topics of best practice books were investigated to get an understanding of what competencies are central for purchasing professionals. Lastly, existing studies focused on on relevant competency areas were reviewed to get an understanding of previous research on the topic.

Based on the responsibilities and objectives of a strategic purchaser, best practices, the existing research on competencies, and the delimitation, hypotheses were built on what competency areas are relevant for strategic purchasing professionals and the contents of these areas. Their interrelation were proposed in order to facilitate the building of competency requirements that are connected to the actual

purchasing activities, as is recommended for competencies in section 3.2.

2.6.2 Case selection: Collect

To get a further insight on what the learning objectives are in higher education purchasing courses, a plan was to make a multiple case study in the topic. The test cases of this multiple case study were the course curricula of purchasing university courses, and the units of analysis were the contents and learning objectives of the courses (Yin, 2009).

The test case courses had to fulfill the following criteria:

- Must be an advanced course in purchasing, procurement or supply chain.
- Must be part of a Bachelor's or Master's program with a focus on industrial engineering.
- Must be a Swedish university course.

The test cases in this study were:

- LTH (Industrial Purchasing, 7.5 credits)
- LiTH (Purchasing, 6 credits)
- Chalmers (Purchasing and strategic supplier collaboration, 7.5 credits)
- Jönköping University (Purchasing, 7.5 credits)
- Luleå University of Technology (Purchase 7.5 credits)

2.6.3 Analyze and conclude

The competency areas chosen in the literature review were validated by comparing with the contents of university courses as well as interviews on the contents that was conducted in conjunction with the content reviews in part 2. This analysis lead to a conclusion of the relevance of the chosen competency areas and their contents. Here, the competency areas themselves were not altered, but the contents were reviewed in order to align more with the experts' opinions.

2.7 Techniques and procedures part 2: Designing a test to measure competencies

Considine et al. (2005) presents a comprehensive guide for designing and validating a MCQ test. The reason why MCQ tests are suitable is presented in section 3.1.2.

While the validity and reliability requirements are presented more in-depth in section 3.3, figure 2.3 shows the overall process for designing and validating the MCQ-test according to Considine et al. (2005).

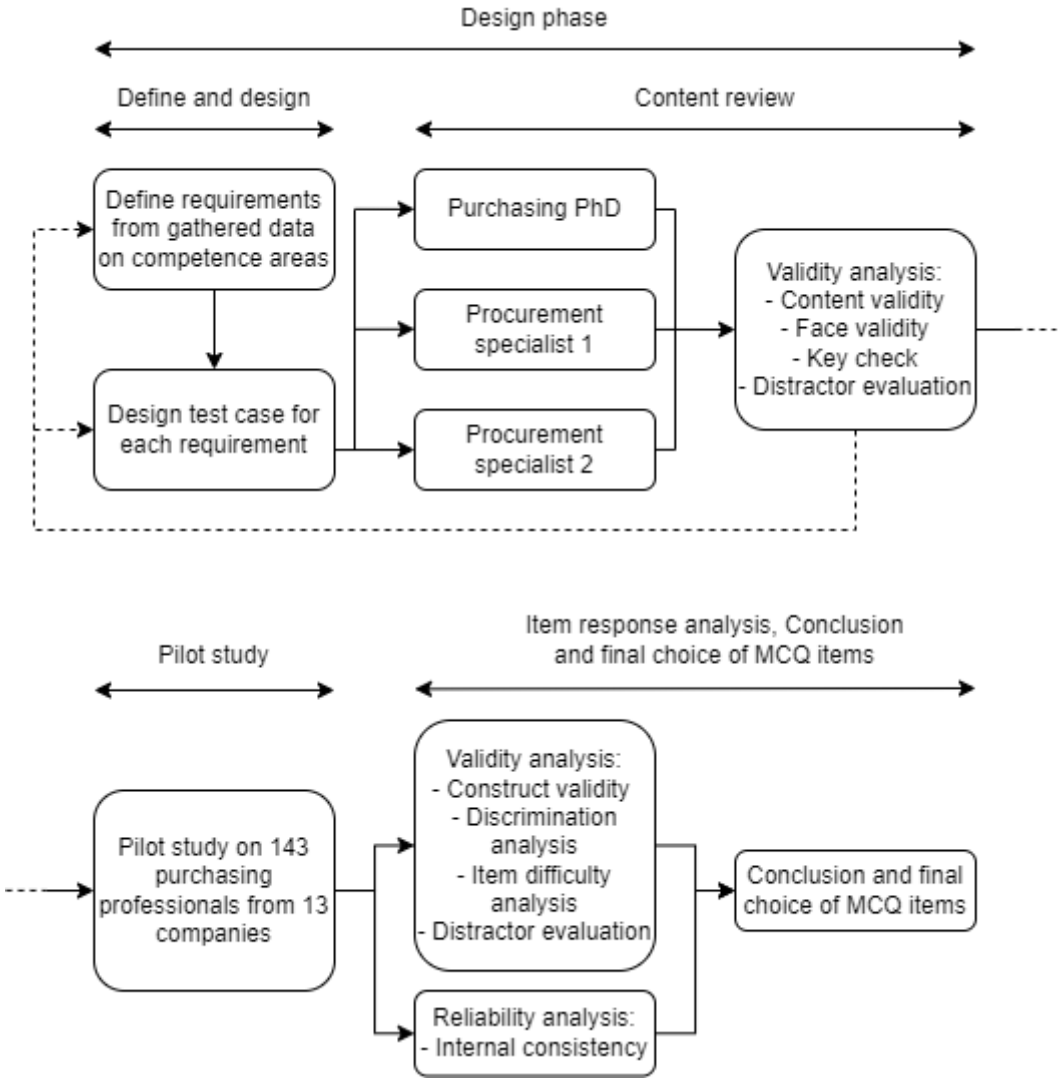


Figure 2.3: Process for design, validation and final choice of MCQ test cases.

2.7.1 Define and design

The methodology began with a design phase leading to a first version of a MCQ test. Based on the information gathered in the first part of this study, a specification of competency requirements was constructed as a framework for testing the contents. More information about the requirement specification is found in the theory chapter 3.3.1. To allow for elimination of questions that are found to be redundant by validity and reliability studies, the number of questions designed were greater than the expected number of questions in the final test (Considine et al., 2005).

Each requirement was constructed be specific enough so it can be tested using one single test case, i.e. one MCQ question. A requirement that for example states “A

strategic purchaser should successfully assess transactional risk" may be tested in several ways. Therefore it is suitable to use requirements both for knowing the purpose of the question as well as being able to alter test case for this or future tests. For each competency area, a set of 11 specific requirements were defined, each covering one important aspect of the competency area. The aim was that testing these requirements would give an overall view of the competency area.

The requirements for each competency area were chosen with an iterative approach. Firstly, a set of candidate requirements were constructed based on the overall theory gathered about the competency area. For each requirement, a test case was designed that fully assesses the requirement. Since the test case fully covered the requirement, these were seen as synonym in discussions. Therefore, only the test case would be addressed from here and not its corresponding requirement. A sample of requirements in the field of Contracting and Contract Management is shown in table .

Table 2.1: Examples of requirements specified for the test case design. In this case for Contracting & Contract Management (CCM)

ID	Subcategory	Requirement (A strategic purchaser shall ...)
CCM1	Pricing methods	... Understand the risks or benefits of fixed / cost based pricing (i.e. incentives)
CCM2	Contact risk	... Understand how to avoid the risk of unforeseen costs. (i.e. cost escalation)
CCM3	Contact risk	... Act correctly when a more riskful alternative is presented (i.e. volume commitment)
CCM4	Contract management	... Know how to avoid repetition when a supplier has not performed
...	...	...

Content review

Content reviews were performed during the design phase and focused on selecting a set of 11 test cases that all passed the qualitative validity requirements, i.e. content validity, face validity, Key check and distractor evaluation (Considine et al., 2005). In order to pass the content validity requirement, the test case itself had to successfully measure the competency in question, and the mix of 11 test cases needed to sufficiently cover the competency area. Although some requirements could only be tested through a pilot study, such as discrimination and difficulty, these were still taken into consideration in the design phase (Considine et al., 2005).

The content reviews were performed separately by the expert panel. Since each test case briefly covered a small and specific topic, and due to the structured and repetitive nature of the questions, a structured interview was suitable in order to cover each necessary question for each test case (Saunders et al., 2009). The interview guide was constructed based on the qualitative validity requirements presented in section 3.3 and is presented in appendix A.

The content review took place during the final stage of the design phase. In this manner, any suggested and accepted improvements were implemented in the design phase and validated once again until the test questions passed the validity requirements from the expert's point of view. If any question was changed in the content review, the answer would be iterated through the other experts to ensure consensus about the new version.

The content reviews were undertaken both during both one session and several sessions. The reason why some of the reviews were conducted in several sessions is because the test areas were finished at different points of time, and it was most efficient to let the purchasing specialists do the reviews for each area whenever they have the time. After the content review was finished and the chosen questions were valid, the design phase of the test was finished and the test could be validated statistically in the pilot study.

2.7.2 Pilot study

Gathering participants for the pilot study

The test was validated on participants in a pilot study and an analysis of the results was conducted in order to get a final version of the test. For this survey based research to be effectively generalizeable, it was important to make sure that the sample was representative for the entire population and also having a large enough sample (Saunders et al., 2009; Considine et al., 2005). With the target group of strategic purchasing professionals defined, the first step was to gather participants.

The sampling method used for selecting participants was voluntary response sampling, meaning that participants were asked to be a part of the tool and the ones who volunteered would make up the sample. The companies were chosen from a convenience sampling method, meaning it included companies most willing to participate. This sampling method was practiced by contacting all medium- and large cap companies in Nasdaq Stockholm by phone and email and requesting to speak to CPO:s or purchasing managers. In some cases, snowball sampling was practiced by purchasing managers referring to other companies (McCombes, 2022). This led for instance to a small number of small cap companies also participating.

As stated in section 1.5, the each participant was assured to conduct the test anonymously. As a result of this sampling method, a total of 143 purchasing professionals from 13 companies were gathered. A table with information about the resulting participating companies is presented in table 2.2.

Table 2.2: Participants in the pilot study

Company ID	Enterprise	Participating regions	Number of participants	Contact person role
Company 1	Large Cap	Sweden, Europe, Asia	51	Global Sourcing Director
Company 2	Large cap	Sweden, Europe, Asia, Americas	30	Group Procurement Director
Company 3	Large cap	Sweden	9	Director Strategic Purchasing
Company 4	Large cap	Sweden	5	Strategic purchasing manager
Company 5	Large cap	Sweden	3	Purchasing Manager
Company 6	Large cap	Sweden	2	Category Manager
Company 7	Large cap	Sweden	2	Supply Chain Manager
Company 8	Large cap	Sweden	1	Strategic Purchaser
Company 9	Medium cap	Sweden, Europe, Asia, Americas	23	EVP Global Sourcing
Company 10	Medium cap	Sweden	5	Vice President Sourcing
Company 11	Medium cap	Sweden	4	Chief Procuremenet Director
Company 12	Small cap	Sweden	6	Purchasing Director
Company 13	Small cap	Sweden	2	Head of Procurement
Total			143	

Gathering validity data for the pilot study

The construct validity is be central the purpose of understanding to which extent the test measures the strategic purchasing competencies as defined in the definition chapter. More on construct validity is presented in section 3.3.2. The construct validity was assessed in the pilot study by testing to what extent the purchasing assessment tool score correlates to a subjective measure of competency based on the of the purchasing managers' subjective experiences of the participants' overall competency.

In order to realize the subjective evaluations, the purchasing managers were asked to evaluate their corresponding participants based on purely strategic and analytic skills and abilities to make correct judgments based on data in a purchasing setting, just as the definition in the study suggests. The scoring was a seven point Likert scale (Joshi et al., 2015) and is presented as:

1. Poor
2. Clearly below average
3. Slightly below average
4. Average
5. Slightly above average

6. Clearly above average
7. Excellent

This way, a hypothesis of the significance of the test result were hypothesized and tested through correlation between the test score and this subjective evaluation value. More information of the choice and setup of this hypothesis is presented in section 3.3.2.

Conducting the pilot study

The test specifications needed to resemble the actual future test in order to be effective (Considine et al., 2005), and these were therefore defined before the pilot study and presented here. The test is based on multiple-choice. For each question, there are five potential answers to choose between. Only one of the answers will be correct, and you can only pick one answer for each question. There is one question at a time and no chance to go back to any previous question. The test will go through seven different areas with 11 questions each. There will be a time limit for each question, 1 min 45 seconds for most questions and 2 min 15 seconds for a few more complex questions. The time limit will reset for each question. Grand total of 2 hour test. Allowed resources are pen, paper, and calculator. The participant is not allowed to search the web or to collaborate. An example test case is shown in figure 2.4.

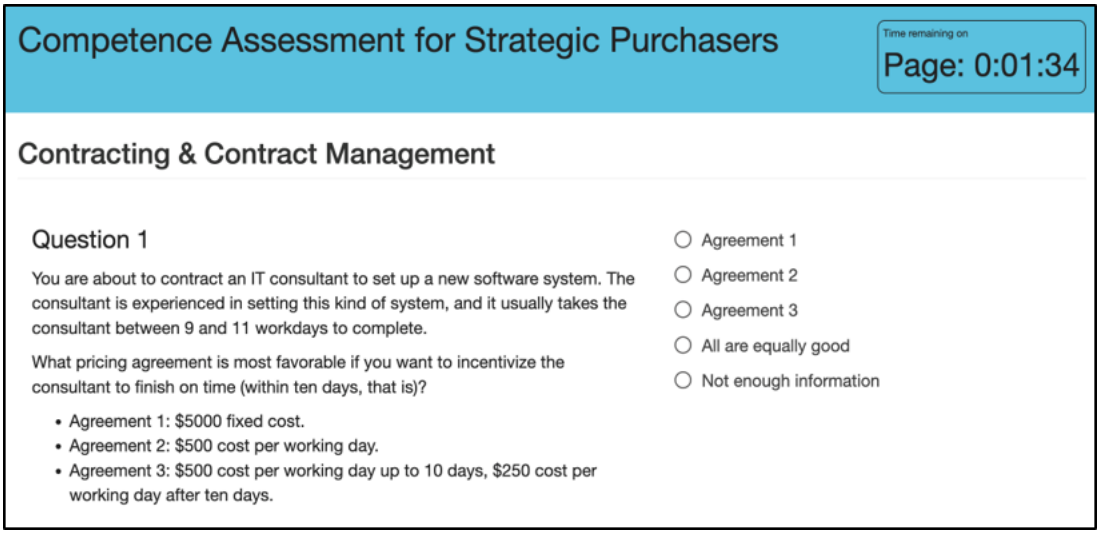


Figure 2.4: Example test case.

2.7.3 Item response analysis, conclusion and final choice of MCQ test cases

After the pilot study, MCQs that did not pass the requirements for discrimination were considered to be replaced by questions with desirable difficulty (Considine et al., 2005). Also, questions with distractors that performed better than the key

were critically evaluated for any ambiguities. The discrimination and difficulty measures are further presented in section 3.3.2.

The test score were also compared with the subjective evaluation of purchasing managers in order to accept or reject the hypothesis that the test fulfils its purpose. This hypothesis is critical in order to validate if the purpose of this study has been fulfilled at all. After this, the reliability of the test was statistically assessed in order to specify in what ways the test can be reliably used.

2.7.4 Final benchmark design and result presentations for participating companies

The benchmark was constructed by creating an empirical distribution of the score for each category. with this benchmark, the each participant could, and may in future uses, be assigned a percentile score compared to the rest of the participants. The resulting benchmarks are presented in section 4.3.2. When the test and validation was complete, the results were presented to the companies as aggregate or individual, in total or area wise. An example of the results that were presented is shown in figure 2.5, where the percentiles are grouped together in quintiles.

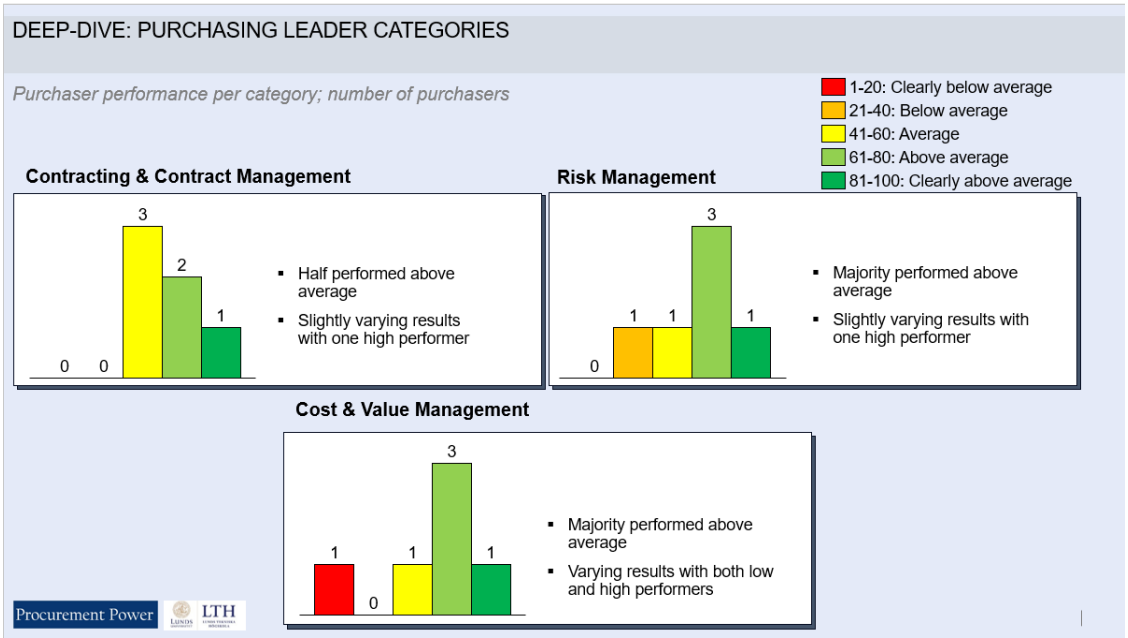


Figure 2.5: Example of the data presented to the companies.

2.8 Credibility

Both the process of identifying competency areas and contents, as well as the design of the MCQ test cases, need to have a focus on credibility. Yin (2009) describes four applicable tactics for credibility, that is, Construct Validity, Internal Validity, External Validity and Reliability. Considine et al. (2005) relates mainly to fulfilling rigorous validity and reliability measures that are further described in

the theory chapter. These are however in similar domain as Yin (2009) and will also be discussed in the following sections.

2.8.1 Construct validity

Construct validity is to what extent the correct measures are taken for analyzing the desired concepts (Yin, 2009; Considine et al., 2005). For the multiple case study, several sources of evidence have been used to both find and validate the tests. First off, the competency areas are proposed using three general sources of information, that is, responsibilities and objectives of a strategic purchaser, best practices, and existing research on competencies. The case studies for validating these first looks at five sources of evidence with similar purpose and high comparability. The competency areas and their contents are further validated through discussions with three experts whereas two are industry specialists and one is from academia, yielding diverse views.

For the test validation, the combination of qualitative and statistical measurements are methodologies of assessing the fit for purpose. The qualitative measurements will be iterated in order to make sure that the test reflects the areas that are supposed to be tested. The fit for purpose will later be assessed using quantitative construct validity measurements from the pilot study, both for the test itself and each of the test cases. These measures will be further described in section 3.3.2.

2.8.2 Internal validity

Internal validity relates to how successfully the chosen measures manage to capture the desired concepts (Yin, 2009). In part 1, the course curricula are chosen from a variety of courses with little limiting criteria. The interviewees are also interviewed separately and in a highly structured manner which reduces the risk of biases.

in part 2, the content validity, face validity, distractor evaluation and key check measures aim to reduce any biases and make sure that the competencies are correctly measured. These will be further described in section 3.3.2.

2.8.3 External validity

The external validity is to what extent the findings can explain the actual concepts (Yin, 2009). The experts assess both if the selected test cases sufficiently cover the competency area in question, as well as if the content area is relevant to a strategic purchaser. Both the data gathering, analysis and conclusion of competency areas in part 1 will also be validated using the test.

For the test validation, the construct validity measure statistically assess the success of the test to fit its purpose for both each of the test cases and the entire test itself. This will be further discussed in section 3.3.2.

2.8.4 Reliability

The reliability is to what extent the study can produce consistent results (Yin, 2009; Considine et al., 2005). For the first part, the choice of test cases as well as the choice of experts aims to create a diverse set of evidence that produces results that should be consistent if reproduced. The theory is also gathered from multiple sources across different domains in order to capture both practical and theoretical points of view. The aim to make unbiased data gathering also contributes to reliable results. Since the interviews are structured and aimed at specific topics, there is somewhat bias towards the topics and therefore the experts might have come up with other competency areas if totally unbiased. However, since the aim of this thesis is to find some relevant areas and not the key areas, it successfully fulfils this purpose.

For the test itself, there are some measures aimed towards assessing the reliability of the test, and in this case the internal consistency statistically evaluates to what extent the results would be consistent in a new pilot study.

3

Theory

The theoretical chapter will be divided into three sections. Firstly, the importance of, and approaches to, measuring competencies will be presented. Secondly, the literature study on competencies necessary for strategic purchasers will be presented and lead to hypothesis of relevant competency areas. The final section will present the requirements for a valid and reliable multiple choice question (MCQ) competency assessment.

3.1 Competence measurements

3.1.1 The connection between purchasing competencies and business performance

In the introduction chapter, a brief presentation of purchasing's link to business performance was presented. The purchasing function's value generation potential is defined by how well the corporate strategy can be trickled down to the purchasing function's operations. The possibility is established by a series of layers that must be aligned (Pohl and Förstl, 2011; Cousins et al., 2008). First of all, the purchasing strategy needs to be aligned with overall corporate strategy, which is enabled through strategic integration of the purchasing function (Pohl and Förstl, 2011; Cousins et al., 2008; Eatough, 2014). The purchasing strategy is implemented through practices and activities which need to be aligned with the purchasing strategy. For this, purchasing performance management systems (PPMS) are important in order to measure the purchasing performance based on the objectives of the purchasing strategy (Pohl and Förstl, 2011; Cousins et al., 2008; Eatough, 2014). The strategic alignment model presented by Cousins et al. (2008) and seen in figure 3.1 summarizes the interrelation between the layers that enable purchasing performance.

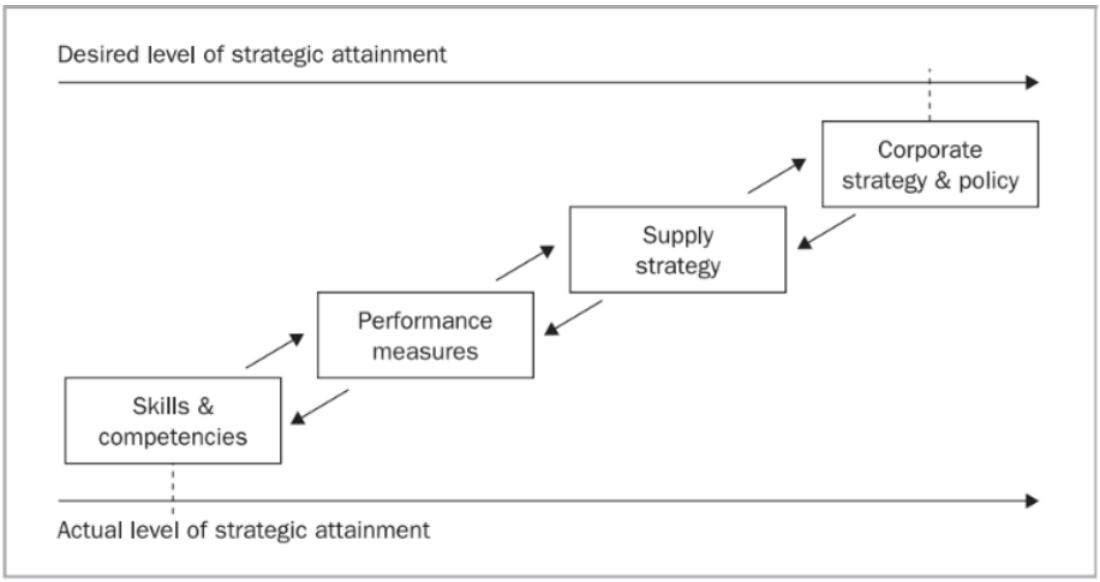


Figure 3.1: Strategic alignment model for purchasing (Cousins et al., 2008)

3.1.2 Reliable approaches to measuring competencies

Conversely to PPMS that are often used for assessment of competency development needs today (Pohl and Förstl, 2011; Aulia and Isvara, 2021), competency assessments are diagnosing the underlying factors that enable a higher performance of the purchasing functions and are therefore more explanatory for this purpose (Deardorff, 2009; Cousins et al., 2008; Eatough, 2014).

There are several ways of measuring competencies for professional development. In the comprehensive handbook of intercultural competence, Deardorff (2009) surveyed a group of 20 leading experts in in the field whether competencies could be assessed interculturally using different measurements. In total, ten different types of measurements were accepted by 80 percent or more of the experts including for example case studies, interviews, analysis of narrative diaries, self-report instruments and observation by others. Two of the quantitative measurements that were accepted were purely quantitative. These were "developing specific indicators for each component/dimension of intercultural competencies and evidence of each indicator" and "triangulation (use of multiple data collection efforts as corroborative evidence for the validity of qualitative research findings)".

In an article in the Human Resources Professional journal, Meger (1996), as cited in Cousins et al. (2008), criticises subjective measures for competence evaluation of individuals and poses ten arguments against it. For example, the results are affected by for example situational factors, behavioural traits, and personal opinions. The evaluators might also have varying perception of the traits that are evaluated, leading to unreliable results and definitions. Meger (1996) underlines the need of establishing objective and fair assessments that are clear and easily understood. Deardorff (2009) also mentions the fact subjective evaluations might require significant the time, money and support from management. In competence development, Kock and Eilström (2009) argues that the effort needed might lead to

companies not performing sufficient pre-study leading to a mismatch in competence development demand and actual competence development implemented.

To conclude, there is evidence suggesting that easily understood, objective, reliable and economically efficient measurements of competencies has several benefits. Reliable quantitative ways of assessing competence include developing indicators specific for an area. The choice of having an multiple choice assessment for seven specific areas will aim to fulfill these objectives. More on the characteristics of MCQ assessments will be discussed in part 2.

3.2 Part 1: Necessary competencies for a modern strategic purchaser

This chapter aims to investigate what competency areas are relevant for a high-performing strategic purchaser. A suitable number of competence areas will be suggested to later be validated in the coming chapters.

There is an important distinction between competency/competencies and competence/competences that is often interchanged in literature. Competencies relate to the skills and knowledge required for being able to perform a certain job in an efficient manner (Cousins et al., 2008), while competences are more output related meaning the ability to carry out a job effectively, and is thus unrelated to the job itself. In order to show competence within a certain job, it is a prerequisite to first acquire the relevant competencies for the job (Cousins et al., 2008). When evaluating competencies, the focus will therefore lie on relating the assessment elements to the job of purchasing and thus relating to the activities of the strategic purchasing professional.

3.2.1 The activities and objectives of a strategic purchaser

The purchasing professional role consists of a diverse set of responsibilities. Figure 3.2 shows the main tasks of the purchasing process. The first process, so-called source-to-contract, has a more strategic focus, while the second and third processes, purchase-to-pay, are more operational (Backstrand et al., 2019). The main responsibilities of a strategic purchaser are included in the source-to-contract process, including Supplier Relationship Management & Performance Management. (Backstrand et al., 2019; Inköpsdesign, 2022).

The objectives of the strategic purchasing function has also expanded from its initial focus on cost reductions. Today's purchaser is part of value creation in a larger sense and needs to balance value cost savings, value improvement, risk management (Van Weele, 2009; Eatough, 2014), and positive CSR impact (Carter and Jennings, 2004).

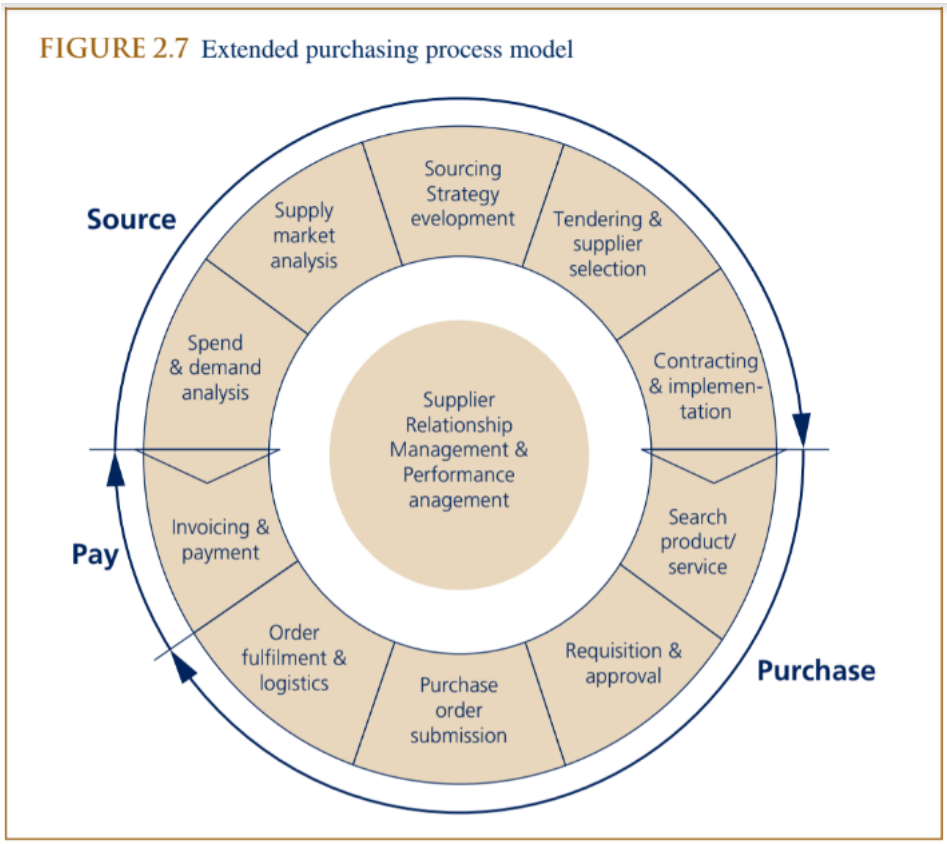


Figure 3.2: Model of the purchasing process

3.2.2 Necessary competencies based on existing research and best-practice guides

There is extensive research on what purchasing competencies are important for successful strategic purchasing, although it tends to show inconsistent results and in some cases vague definitions. Giunipero and Pearcy (2000) mentions in their article "World-Class Purchasing Skills: An Empirical Investigation" seven skills which are strategic skills, process management skills, team skills, decision making skills, behavioral skills, negotiation skills and quantitative skills. Handfield et al. (2015) talks about several relevant skills such as team building, strategic planning, effective communication and relationship management. Others are technical skills, broader financial skills, and legal skills. Tassabehji and Moorhouse (2008) provide a comprehensive picture of qualifications. These are grouped into technical skills, interpersonal skills, internal and external enterprise skills, and strategic business skills. The Danish Purchasing and Logistics Forum (DILF, Accessed 2022) mentions Strategic Sourcing, Negotiation & Contracting, Cost Value Management, Supplier Management, and Risk Management as the key functional competencies for strategic purchasers. Bals et al. (2019) describes that qualifications for best practice strategic purchasing includes communication, relationship management, and strategic and analytical thinking.

The competencies are also argued to vary depending on the category the purchaser might be involved with. Cousins et al. (2008) relates competencies to the Kraljic matrix, meaning that necessary competencies may differ depending on type of

product or service. For example, risk assessment and relationship management being more important for items with a more complex and riskful supply market, and negotiation skills and cost management being more important for less complex and riskful supply markets.

3.2.3 Trends towards higher and more specialized education for strategic purchasers

As the role of strategic purchasing professionals evolve, the qualifications for the role evolves as well, and the demand for excellent purchasers has increased (Carr and Smeltzer, 2000; Giunipero et al., 2006). As a consequence, higher education institutions have experienced a need of incorporating purchasing and supply chain management in MBA and MBA/Tech education (Carr and Smeltzer, 2000; Giunipero et al., 2006). In the empirical study chapter, it will be shown that several higher education institutions in Sweden have incorporated purchasing courses as part of their curricula. The contents of these areas will be further investigate in order to validate the relevance of the hypothesized competency areas.

3.2.4 Hypothesis of relevant purchasing competency areas and their contents

Based on the responsibilities and objectives of a strategic purchaser, the existing research and best-practice-guides on competencies, and the delimitation for hard and functional competencies, a deep-dive of seven competency areas that are hypothesized as relevant are presented.

The first three competency areas are Sourcing Strategy, Negotiation, and Contracting & Contract Management, covering the relevant activities for strategic purchasing professionals. The next three are Cost & Value management, Risk Management and Sustainability, covering the value creation objectives that are present in all activities. The final is financing, based on the fact that quantitative and enterprise skills is a prerequisite for purchasing courses. For the test to be valid, the test areas and items need to pass both content validity and construct validity measures. Therefore, these hypotheses are set up and must be tested. The description as well as hypothesis for the chosen competencies are presented below.

Sourcing Strategy

Sourcing strategy in this case will consist of strategic sourcing as well as supply market analysis. Strategic sourcing involves managing the supply side in as effective manner as possible, and is relevant for a successful supply chain (Talluri and Narasimhan, 2004). It involves both supplier selection as well as supplier development in order to select the right suppliers for long-term partnerships (Talluri and Narasimhan, 2004).

Supplier selection involves assessing suppliers through supplier audits, quotations, cost modeling, financial analysis etc (Van Weele, 2009). Supplier development encompasses optimizing the existing and future supplier relationships, through identifying and analyzing the current situation and identifying practical improvement initiatives (Van Weele, 2009). Supply market analysis can be analyzed both with qualitative and quantitative data includes supplier segmentation such as the Kraljic matrix, competitiveness, market risks, market trends, etc. (Department of Housing and Public Works, 2010; Webb, 2017; Van Weele, 2009).

- H1: The test category Sourcing Strategy is valid.

  H1a: Sourcing Strategy is a relevant competency area for strategic purchasers.

  H1b: Sourcing Strategy will successfully be measured by the test.

Negotiation

At least 20 percent of a purchaser's time is taken up by preparation for, and the actual, negotiation (Van Weele, 2009). Collecting and analyzing the right data for negotiation allows the purchaser to reach agreements on reaching, for example, better rejection rates, less lead time and cost reductions (Van Weele, 2009). A negotiation method should reach an agreement that is sustainable and fair, it should be efficient and it should leave the relationship in at least the same condition as before. (Fisher et al., 2011). The negotiation stage will result in a contract, and therefore they are complementary (Van Weele, 2009). However, they will be treated as two different areas due to their complexity.

The book 'Getting to Yes' is one of the most recognized books for negotiators in the world of management and sales. In the book, Fisher et al. (2011) describe in the book how to conduct a principled and efficient negotiation and centers it around the People, Interests, Criteria and Options (PICO) principle. People means that the parties should separate themselves from the problem and tackle it together. Interests means to identify the shared and the conflicting interests. Conflict of interest needs to be dealt with in negotiation. This is further mentioned by Van Weele (2009). Criteria means making sure the result will be based on objective criteria, coming from fair standards or a fair procedure. To do this, the issue should be framed into appropriate and objective criteria that are legitimate, practical and independent of the will of each side. Options means to find a variety of options that can be successful, and in that way create a room to look for mutual gains. If there are conflicting interests, for example dovetailing is a good alternative where you look for solutions that are low cost for you and high value for them, and vice versa.

There are other aspects to negotiation such as evaluating the negotiation position of each party (Van Weele, 2009), and improving the position by for example finding the best alternative to no deal (BATNA), bringing new suppliers to the market, change purchasing volumes, or consolidate orders (Fisher et al., 2011; Paranikas et al., 2015).

- H2: The test category Negotiation is valid.

  H2a: Negotiation is a relevant competency area for strategic purchasers.

  H2b: Negotiation will successfully be measured by the test.

Contracting and Contract Management

The contracting procedure is divided into three phases, according to Van Weele (2009): pre-contractual, contract negotiation and closure (contracting), and post-contractual (contract management). The different phases are all necessary for the contracting procedure to be successful and cannot be evaluated independently. (Van Weele, 2009). Contracts are the link between companies in a value chain, and these have become more complex and varied, with risks such as goal incongruence and information asymmetry needed to be dealt with efficiently (Van Weele, 2009).

Lowe (2007) defines three important aspects for successful contracting and contract management. The first is defining the parties' responsibilities, objectives, and priorities, including the expected performance and time frames. The second is assigning responsibility for all risks, including subcontracting risks, investment risks, and delay charges. The third is determining appropriate payment conditions for incentivization and cost escalation reduction. The last can be dealt with by having for example outcome-related contracts, aligned goals or information systems for monitoring progress Van Weele (2009).

- H3: The test category Contracting & Contract Management is valid.

  H3a: Contracting & Contract Management is a relevant competency area for strategic purchasers.

  H3b: Contracting & Contract Management will successfully be measured by the test.

Cost and Value management

Cost and value management has become more complex in the increasingly competitive world, including factors such as Total Cost of Ownership (TCO) optimization, cost avoidance, and value creation for long-term profitability. (GLOMACS, 2022). Strategic cost management includes of identifying all expenses, cost drivers, and cost-cutting or cost-reduction initiatives across the supply chain with a TCO perspective (Van Weele, 2009; Ellram, 1993). Value analysis or value engineering involves studying all elements of cost and suggesting improvements while keeping the function of the product or service (Kendt and Nichols, Accessed 2022).

- H4: The test category Cost & Value management is valid.

  H4a: Cost & Value management is a relevant competency area for strategic purchasers.

H4b: Cost & Value management will successfully be measured by the test.

Risk Management

Risk Management involves to identify, prioritize and act upon risks in the purchasing process to ensure effective management of supply chains (Hallikas and Lintukangas, 2016). Securing an undisrupted flow of supplies is one of the main responsibilities of purchasing professionals (Hallikas and Lintukangas, 2016; Ellis et al., 2010). With supply chains becoming more complex and uncertain, maintaining this is becoming an increasingly harder task and companies are therefore actively engaging in reducing these risks. (Hallikas and Lintukangas, 2016; Ellis et al., 2010).

Smeltzer and Siferd (2006) summarizes risks in all stages of the purchasing process. For example, internal demand might be realized too early or too late, the requests for quotations (RFQ:s) might be overspecified, underspecified or misaligned, and the chosen supplier might not be capable or incentivized to perform according to specifications with the terms negotiated. For the resulting agreement with the supplier, Van Weele (2009) divides it up into three types of risk. Technical risks relates to the functionality or performance if the supplier. Commercial risks is related to cost deviations. Contractual risks and Performance risks is related to the capability of supplier working in the scope of what it was hired for (Van Weele, 2009). Risk in a purchasing setting is in congruence with the probability and magnitude of the risk, and risk management includes prioritizing and mitigating both these dimensions (Ellis et al., 2010). In a changing environment, the alterations in power balance between parties is also relevant to sustainably maintain a low risk (Van Weele, 2009).

- H5: The test category Risk Management is valid.

  H5a: Risk Management is a relevant competency area for strategic purchasers.

  H5b: Risk Management will successfully be measured by the test.

Sustainability

Sustainability in the context of purchasing has several linked terms, such as Socially Responsible Procurement (SRP), Environmental purchasing and supplier management (EPSM), Purchasing social responsibility (PSR) and it is overall aimed towards improving supply chain social and environmental performance (CIPS, Accessed 2022; Tate et al., 2012; Carter, 2005).

Many companies are in early stages of development towards sustainable purchasing (Tate et al., 2012). However, measurements of purchasing performance in a sustainability dimension is more developed (CIPS, Accessed 2022). Purchasing has many opportunities of impactful improvements (Tate et al., 2012) and the topic is

believed to be increasingly relevant in the future (Van Weele, 2009). In addition to being the ethically correct approach, there is a connection between sustainable purchasing and supplier performance, inducing further incentives to drive towards sustainable purchasing (Carter, 2005).

Prerequisites for driving sustainable work includes understanding of resource-based views, logistics, corporate social responsibility, and circular economics, social and health implications, and local economy (Carter, 2005; Kiss et al., 2019). Sustainable purchasing can be driven throughout the purchasing process, including transportation, means of production, and supplier relationships (Kiss et al., 2019). The purchasing professional is also the relevant player in assessing to what extent the suppliers are complying to, and following sustainability requirements (Van Weele, 2009). Ways of driving sustainable work for purchasers is to have a Supplier Sustainability Codes of Conduct signed by the supplier, based on for example United Nations Global Compact Principles (UNGCP), or Global Reporting Initiative (GRI) (Van Weele, 2009).

- H6: The test category Sustainability is valid.

  H6a: Sustainability is a relevant competency area for strategic purchasers.

  H6b: Sustainability will successfully be measured by the test.

Financing

In this study, financing and finance is used interchangeably. Although financing only means the actual funding of a project and thus is a subgroup of finance, it is often regarded as synonyms (AE University, Accessed 2022). The reason is because of the communication purpose, where financing was assumed to be more relatable for strategic purchasers and finance might be perceived as too wide and unnecessary.

Based on much of the earlier competencies, it shows evidence that the strategic purchaser's role incorporates quantitative and financial analysis in several activities, for example profitability, budgeting and cash management (Van Weele, 2009). Good understanding of financing enables decision-making in allignment with corporate goals, better communication between purchasing and other business functions, and sustainable cost- and risk benefits (Medius, 2022; Van Weele, 2009). Financing is an important cornerstone for quantitative concepts such quantitative supply market analysis, vendor rating for supplier assessment, financial supplier assessment, budgeting, cost optimizations (Van Weele, 2009; Medius, 2022).

- H7: The test category financing is valid.

  H7a: Financing is a relevant competency area for strategic purchasers.

  H7b: Financing will successfully be measured by the test.

3.2.5 Conceptual framework for the selection of test requirements

All chosen areas except financing is either an activity or an objective. There is a conceptual distinction between the two, where objectives represent the desired outcomes, whereas activities are a collection of tasks that achieve an objective (University of Texas School of Information, Accessed 2022), financing on the other hand is a prerequisite skill set that lays ground for succeeding in an activity (Vuong et al., 2011). Based on this, a conceptual framework is formed that will form a basis for the choice of comprehensive and meaningful test requirements that tests job related competencies and not competence. This framework is presented in figure 3.3.

		Activities		
		Sourcing Strategy	Negotiation	Contracting & Contract Management
Objectives	Cost & Value Management	e.g. supply base reduction	e.g. bargaining power	e.g. cost escalation
	Risk Management	e.g. supply risk reduction	e.g. negotiating performance measures	e.g. forex risk
	Sustainability	e.g. sustainable sourcing	e.g. sustainable negotiation objectives	e.g. codes of conduct
Foundation	Financing	e.g. financial supplier assessment	e.g. profit margin impact	e.g. cashflow impact on payment terms

Figure 3.3: Conceptual framework for the choice of comprehensive test requirements.

This framework is not meant to be mutually exclusive, since some requirements might cover more than one of the elements. Each requirement in a given area should however require to be traceable to one of its corresponding rows or columns in order to be meaningful, and all elements in the corresponding row or column should be covered in order for the area to be comprehensively tested.

3.3 Part 2: Measuring competencies using a multiple choice question assessment

Among quantitative measurements for mapping competencies, multiple choice question assessments is one of the most well-established within several domains (Brady, 2005; Tangianu et al., 2018; Buckles and Siegfried, 2006; Walstad and Becker, 1994). Although some argue that there is little to no difference between MCQ assessments and other forms of assessment for economic related competencies (Walstad and Becker, 1994) there are some limitations on complex questions requiring several steps of reasoning (Buckles and Siegfried, 2006). However, it is considered efficient and reliable compared to subjective measurements, based on the requirement of being designed and validated in a highly rigorous manner (Brady, 2005; Tangianu et al., 2018).

This section will first briefly present the theory of requirements as a basis for testing. It will afterwards present what validity Multiple Choice Questions (MCQs) themselves must meet, and the validity and reliability requirements the test and the tested areas should meet. Validity is to what extent the test or test case measures what is to be measured whereas reliability is to what extent the test provides results with as much correlation and as little variance as possible (Considine et al., 2005). Validity is closely connected to reliability, but even if all the validity measures are passed, it is still highly relevant to understand how reliable the measurement is. For example, a scale might prove to measure the weight of certain items successfully. However, the scale might still have vastly different reliability, for example 0.1 percent or 20 percent margin of error.

Considine et al. (2005) presents a comprehensive guide for designing, validating and testing MCQs for educational purpose that will lie as a basis for this section. It should be mentioned that the source is a review article, meaning that is reviewing and synthesizing earlier literature. This source hence provides mostly secondary information and the large amount of primary sources that lie as a basis for the statements are not mentioned in this section but can be found in the source.

3.3.1 Quality assurance by specifying requirements and test cases

Quality assurance using requirements specifications is usually done with software, but has more applications, for example in certifications where requirements lie as a basis for ensuring a certain quality of a product or service (just as this study is doing). (American Society for Quality, 2022; Borg et al., 2022). The idea is that the same quality requirement can be tested in several different ways, so the method is relevant as a basis for course and competence assessments, where the learning objectives of the test should be clearly specified regardless of assessment type (Meger, 1996; Waina, 1969). Unlike course contents, the learning objectives are based on the knowledge that should be inherited by a strategic purchasing professional according to literature.

3.3.2 Validity

Validity is to what extent the test measures what is to be measured. MCQs should be tested using several measurements of validity in order to test validity and to find sources of bias (Considine et al., 2005). Sources of bias could be due to difficult words, bad instructions, unclear statements, too dietary time limit, incorrect construction, too few questions or patterns (Considine et al., 2005). There are three main legs for analyzing validity, namely content validity, face validity and construct validity, see figure 3.4 (Considine et al., 2005). These measures are in line with the credibility tactics presended in section 2.8. The validity of each of these elements must be determined to determine the overall validity of each MCQ in the test (Considine et al., 2005).

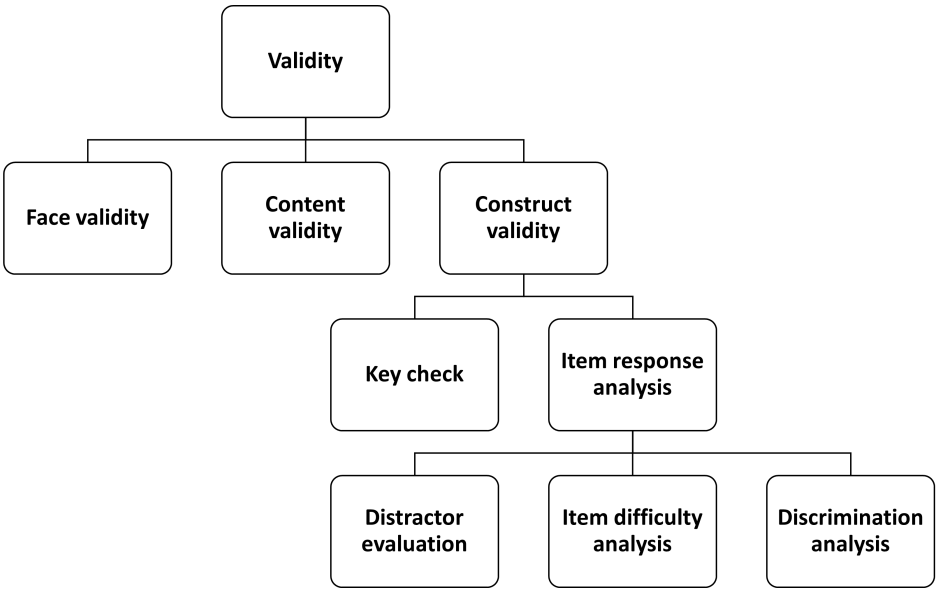


Figure 3.4: Structure for validity.

Face validity

Face validity is to what extent the test has an 'appearance' that is suitable, such as the test clarity, readability and ease of administration (Considine et al., 2005; Wadi et al., 2014). If face validity is not passed, there might be incorrect answers caused due to misunderstanding the question rather than not knowing the answer (Considine et al., 2005; Wadi et al., 2014). This measure is in line with the desired internal validity tactic presented in section 2.8. Face validity is measured through editorial review of grammar, content etc, and analyzed in the pilot study (Considine et al., 2005; Wadi et al., 2014). There are other things to keep in mind such as having simple language, vertical formatting of options, logical and unrelated order, and testing only one attribute (Considine et al., 2005).

Content validity

Content validity is to what extent the content of the test is relevant, suitable and representative of the attribute that is supposed to be tested (Considine et al., 2005), as illustrated in figure 3.5. If content validity is not passed, the test is not measuring everything that is supposed to be measured, or even measuring the wrong things (Considine et al., 2005; Wadi et al., 2014). This measure is also in line with the desired internal validity tactic presented in section 2.8. There is no consensus of an objective way to investigate this but it is usually done through content review, which should be performed by at least three experts in the field being reviewed and who also have some expertise in test tool development (Considine et al., 2005).

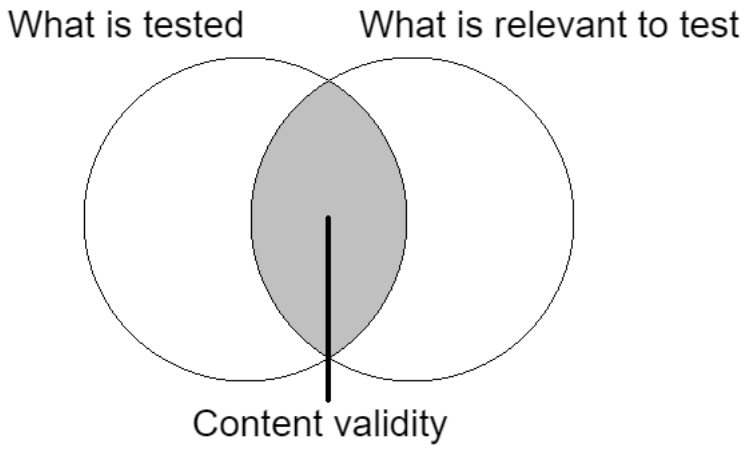


Figure 3.5: Illustration of content validity.

Construct validity

Construct validity is to what extent the test is measuring the theoretically defined attribute that is supposed to be measured, i.e. the more you have a certain attribute, the better you should score on the test (Considine et al., 2005). If construct validity is not passed, the test does not manage to measure the attribute that is supposed to be measured (Considine et al., 2005). This metric corresponds to the desired construct validity strategy described in section 2.8. This validity aspect should be tested using Key check and response analysis with its subgroups item discrimination analysis, item difficulty analysis and distractor evaluation (Considine et al., 2005; Wadi et al., 2014).

According to Fink (2010), are two common ways of assessing construct validity. The first way is to analyze to what extent two constructs that should be similar relate to each other. This test is implemented by hypothesizing that two constructs should correlate and then analyzing the correlation through i.e. a pilot study. The other way of assessing construct validity is by hypothesizing that a tested variable should be able to distinct the outcome of two groups separated by a certain tested variable.

As presented in the methodology chapter, the ability of the test to measure the actual competency as defined in the definition chapter is tested using two separate construct validity measures and set up as a hypothesis:

- H8: The test will successfully measure the strategic and analytic skills and abilities to make correct judgments based on data in a purchasing setting.

  H8a: The correlation between the test score and the subjectively evaluated purchasing competencies will be significantly positive.

  H8b: The correlation between the test score and the subjectively evaluated purchasing competencies will be positive for a significant proportion of companies.

For the first hypothesis, the null hypothesis of is set up so that the Pearson's correlation coefficient $r \leq 0$. The first test variable, the Pearson's correlation coefficient, will follow a Student's t-distribution (Forthofer et al., 2007) and will be tested using the p-value constructed from the formula $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ where n is the number of participants. For the second hypothesis, the test variable follows the binomial distribution and the null hypothesis of a probability of 0.5 of the test showing positive correlation for a certain company will be tested.

Key check

Key check is checking whether the test only has one distinctly correct answer, and not several possible, and it should be conducted by experts in the content area (Considine et al., 2005). If they do not reach a consensus about an MCQ, the MCQ should be revised until there is consensus (Considine et al., 2005; Wadi et al., 2014). This measure relates to the external validity strategy that was discussed in part 2.8, with objective to show whether the test can explain the concepts with little bias.

Item response analysis

Item response analysis is the leg of validity analysis that is checked by numerical analysis of the answer distribution from the pilot study (Considine et al., 2005), and is a subgroup of the validity structure shown in figure 3.4. The analysis consists of three measurements - distractor evaluation, item difficulty analysis and discrimination analysis (Considine et al., 2005). In conjunction with the

Distractor evaluation

A good distractor is one that is selected by those who perform poorly and is ignored by those who perform well, while distractor that is chosen by a small proportion of participants is obviously ineffective and should be replaced (Considine et al., 2005). If a distractor is chosen by more than the correct answer it might indicate ambiguous instructions, a misleading question or some other bias

(Considine et al., 2005). Distractors are evaluated by analyzing the distribution of answers in the pilot study (Considine et al., 2005). In this study, this will be dealt with in the content review in order to work proactively but also assessed retroactively by looking at the response distribution.

Item difficulty analysis

It is important to have an even difficulty distribution to avoid having the test being biased towards a certain difficulty since a test with evenly distributed difficulty can more successfully discriminate between all levels of competency (Considine et al., 2005). For example, if all MCQs have a difficulty of 40 percent meaning that the top 60 percentile are likely to pass the question, then all participants in the 70-100 percentile might score just as good. A higher item difficulty variation will lead to decreased precision, but is important due to aforementioned reasons (Considine et al., 2005). Item difficulty is the percentage of participants who selected the correct answer for that question, and is measured from the answers in the pilot study (Considine et al., 2005). Since a high difficulty may indicate either a tough question or large bias, a high difficulty is dependent on high levels of discrimination (Considine et al., 2005).

Item discrimination analysis

Item discrimination analysis tests how each single MCQ is related to overall test performance (Considine et al., 2005). A question is highly discriminatory if the total result for those who answered correctly is significantly higher than for those who answered incorrectly (Considine et al., 2005). Questions with high degree of discrimination are considered the best because they successfully test the knowledge without having biases such as ambiguity or difficult words affecting the performance (Considine et al., 2005). However, too high degree discrimination can be considered overlapping already tested knowledge seem redundant (Considine et al., 2005). The analysis is made using the data from the pilot study (Considine et al., 2005). For measurement, 'Item to total correlations' is recommended to statistically establish the correlation between the MCQ and the total score (Considine et al., 2005). The item-to-total correlation of 0,25 or higher is acceptable. A correlation of 0,70 or greater means that the question may be redundant (Considine et al., 2005).

3.3.3 Reliability

Reliability is to what extent the test produces consistent scores during repeated measurements (Considine et al., 2005). Compared to for example an essay test where the scoring depends on who corrects it, the handwriting of the participant etc, an MCQ test is generally considered reliable due to its objectivity (Considine et al., 2005). Reliability relates to the measurement concepts of consistency, precision, stability, equivalence and internal consistency and all reliability

measurements are quantitatively measured (Considine et al., 2005). The reliability measures are presented in figure 3.6.

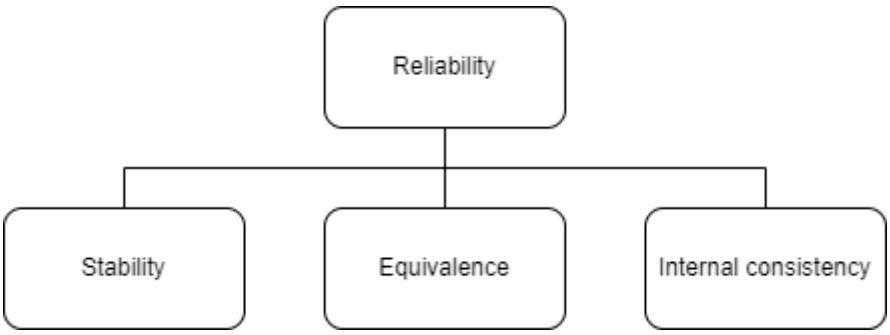


Figure 3.6: Structure for reliability.

Stability

Stability is to what extent the test scores correlate between test and re-test (Considine et al., 2005). This is important when wanting to test an intervention and therefore want the change in test score to be because of reasons not related to the test design (Considine et al., 2005). Since the time frame of this study is cross-sectional and only one pilot study will be conducted, this test will be omitted from this study.

Equivalence

Equivalence is to what extent the test scores correlate between two test sets and is measured by having the same participants taking two tests in succession (Considine et al., 2005). Since only one test will be conducted for each participant, this is not applicable.

Internal consistency

Internal consistency is measuring the reliability of the measurement (Considine et al., 2005). It assesses to what extent the test should produce consistent results by looking at the reliability of the contribution of each test case (Considine et al., 2005). Since this measure relates to the consistency of results, this measure is in line with the reliability tactic that was discussed in section 2.8. There are several ways of measuring internal consistency, and Kuder-Richardson coefficient (KR-20) is considered overestimating the variance of especially large or small results (Anselmi et al., 2019). However, especially for studies with binary results, the KR-20 is suitable to measure the internal consistency. (Considine et al., 2005; Barrett, 2007; Anselmi et al., 2019) The equation for the KR-20 coefficient is calculated as

$$\mathrm{KR}_{20} = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} p_i q_i}{s^2}\right)$$

where k = the number of questions, p_i = proportion answering correctly on question k, and q_i = proportion answering incorrectly on question k (Barrett, 2007).

The KR-20 value ranges from 0 to 1, and a higher value indicates a more reliable measurement (Barrett, 2007). There are no statistical measures indicating what is considered reliable enough, but a general rule of thumb is that a KR-20 value of 0,7 or higher is considered acceptable, and for tests with more than 50 questions the KR-20 value should be greater than 0.8 (Salkind, 2010). Since the internal consistency does not have a statistically defined lowest value for significance, the internal consistency will not be an element of accepting or rejecting the hypotheses of whether the test or any category successfully measures overall competency or a competency area. However, it will be an integral part of the discussion of to what level of reliability it measures the competency for an individual or a group.

4

Empirical Study

4.1 Course curriculum investigation

First, it was discovered that the main textbook for was the same for all universities, namely editions of Van Weele (2009). The comparison was thus made so that the content was examined overall by looking at course elements that were included in the purchasing courses. If the course elements were not specified in the course curricula, the contents of the lectures was examined instead. All course curricula were detailed enough to get an overview of the course contents and objectives. However, the course curricula were written in different levels of detail, with some explaining more in-depth about the course contents and learning objectives than others. The course curricula are not attached in the thesis of in the bibliography chapter. Instead, they are viewed as empirical evidence and a summary of the contents are presented in written form.

To begin with, the activities in the purchasing process were investigated in the curricula. **Purchasing strategies** was explicitly included in all five curricula. This includes supplier market analysis, product group analysis, supplier evaluation and assessment, supply strategies and supplier development (LTH, LiTH, Luleå, Chalmers, Jönköping University). **Negotiation** was included in three out of five course curricula with phrasing such as convincing and arguing, negotiation strategies, and negotiation preparation and negotiation per se (LiTH, LTH, Jönköping University, Chalmers). **Supplier relationship management** was included in five out of five curricula. These included Supplier relations, supplier collaboration, supplier contract terms, as well as supplier relations relationship with environment and ethics (Jönköping University, Chalmers, LiTH, Luleå, LTH).

Secondly, the objectives were investigated. **Sustainability** was present in four out of five course curricula, with phrasings such as as Sustainable purchasing, social, ethcial and environmental bottom line, sustainability aspects, and environment and ethics (LiTH, Chalmers, Jönköping University, LTH). **Cost management** was explicitly included in three out of five course curricula, focusing on cost optimization and total cost of ownership analysis, as well as effective material flows (Luleå, LTH, Chalmers, LiTH). **Risk management** was explicitly mentioned in some form in two out of five curricula, as supply disruption risks and risks

connected to material flow planning. Other course curricula included critically evaluating to purchasing practices and situations (Chalmers, Jönköping University, LTH).

Finally, **Financing** was mentioned in one out of five course curricula (Chalmers). However, the basic course in Industrial Economics was a presequisite for two courses (LTH, LiTH), logistics was presequisite for four out of five (LTH, LiTH, Jönköping University, Luleå), and all courses are advanced level and therefore basic qualifications for advanced level was needed in general (LTH, LiTH, Jönköping University, Luleå, Chalmers).

The general topic of **understanding the role of the purchasing function** to a company's operations was explicitly included in all five course curricula. **Public procurement** was explicitly included in three (LTH, LiTH, Chalmers), **Electronic procurement** was present in two (LTH, Chalmers), and **purchasing law** was present in two (LTH, Luleå). Other than that, there were no other elements that that could not be grouped into any of the other topics.

4.2 Content reviews

Data were gathered during two parts of the purchasing assessment tool. Firstly qualitative data was gathered during the design phase through content reviews, and secondly quantitative data was gathered through the pilot study.

4.2.1 Content review with Purchasing Specialist 1

The first content review was performed through estimated 10 hours of discussions over 10 meetings with a purchasing specialist with vast experience in category optimizations and purchasing competence development. The content review with this specialist was conducted during several sessions. In this manner, any suggested improvements were inserted in the design and validated once again until the test questions passed the validity requirements from the specialist point of view.

Content validity

The chosen competency areas were considered relevant in general. There were however several opinions around the contents of each area. To begin with, there were some requirements that the specialist found had lower priority than others. For example, one of the Contracting & Contract Management requirements stated that the strategic purchaser should understand residual listing. The requirement was found too niche and only relevant for a small proportion of purchasing professionals. One sustainability requirement was to be able to calculate financial consequences of emission rights. This requirement was not found relevant for a strategic purchaser and was therefore omitted.

The specialist was surveyed if the mix of test cases was sufficient for covering each

competency area comprehensively. There were some requirements missing for each area, and these were inserted. For example, the risk management area should include understanding what supplier relations should be brought to the steering group's attention during an investment. Another example is that Contracting & Contract Management should include understanding what relevant contractual terms to include if the supplier will leverage multiple sub-suppliers in development and maintenance of a service.

Face validity

During this initial content review of the design process, there were several points that needed to be clarified. For example in Sourcing Strategy, one test case is to understand what guiding principles to follow to ensure material supply when transitioning to Just in Time. Initially, the question was formulated as what relationship to have with the supplier, which can have an ambiguous connotation.

Key check

In key check, some comments were raised. This was mainly due to ambiguous wordings. For example, the relevant "If you can tender the services centrally to capture volume synergies" was initially formulated as "if you can use the same supplier for several services" which does not have the same meaning.

Distractor evaluation

Distractor evaluation was the section with most comments. The relevant insight from interviews with the specialist was that distractors in particular was a difficult subject without extensive experience. It was therefore an important asset to have several discussions around relevant and difficult distractors that were distinctively incorrect but plausible without extensive experience. For example, in negotiation, one question asks what to do if a supplier is not competitive in price. The distractor "Accept the difference, the supplier has more expertise within its own market and we should build relationships on trust" was proposed by the specialist and later used.

4.2.2 Content review with purchasing specialist 2

The second content review was performed through estimated 5 hours of discussions over 5 meetings with a purchasing specialist with vast experience in category optimizations in several industries.

Content validity

According to the purchasing specialist, the choice of competency areas were all relevant for a best practice strategic purchasers. There were no specific competency areas that were more relevant, although some of the areas could be altered content wise and some examples are highlighted. Contracting & Contract Management should include understanding of price adjustments in a contract, since it can easily escalate in cost. This test case was later added. Another case that was added was in the test was a question about financial risk, e.g. parent company guarantee. In Contracting & Contract Management, the test case of understanding incoterms was considered too specific, leading to the removal of that test case.

Face validity

There were some comments about the setting of certain questions. One of the questions is about analyzing request for quotation (RFQ) data from four different suppliers and understanding what is a relevant action before the upcoming supplier negotiations. This was initially formulated as being an analysis to perform during a negotiation, which is an unlikely thing to do at that time.

Some comments were also made about the length of certain alternatives. Some groups of alternatives were so long that it was a prominent risk that the participants would not be able to finish reading the questions during the course of the time limit. An example is in a negotiation question where the alternatives were written in the form "Focus on the price per call compared to the average time to resolve issue, and negotiate with the service with the highest profit per call". These alternatives were reduced in length to only "Focus on the price per call compared to the average time to resolve issue".

Key check

There were some comments regarding the language use for certain alternatives. An example is in the question "When is it most suitable to negotiate a lower price with an existing supplier?" where all five alternatives needed clearer formulations. Although the alternatives were valid, the ambiguity of the formulations could be misleading.

Distractor evaluation

There were comments about a few distractors. Foremost, in one question in Contracting & Contract Management, the participant is asked to understand the risks of having more favorable contractual conditions from a supplier compared to other customers. His comment was that the correct answer "The supplier might less prioritize your orders in case of a shortage" might depend on the kind of supplier.

4.2.3 Content review with Purchasing PhD

The content review was conducted as an interview that lasted five and a half hours with a purchasing PhD at Lund University with experience in exam design. Each question, area and the test in its entirety were treated according to the interview guide found in Appendix A.

Content validity

For all seven areas, the overall comment was that all tested areas were relevant. In the course exam, it is usually a question of weighted point evaluation system. Financing and risk management is not really included in the course content right now, although financing has actually been requested by some students. Cash flow is important, among other things, in construction projects. For other companies, it is not as important.

For each area, there was a good mix of questions. The negotiation area could use some question of principle negotiation itself, and Contracting & contract management could use some basic question about the most important contents of a contract. In sustainability, one questions were a little bit too specific. Something else that can be included is to connect to Van Weele's purchasing process, such as "How can the purchaser make a more sustainable purchasing in the three first parts of the Van Weele process? "

Face validity

In the face validity area, there some cases where the test item is not relevant to the area. One item in Sustainability is asking what planetary boundaries are mainly affected by carbon dioxide emissions. This is a little bit too specific question maybe. Once contracting & contract management question deals with price adjustments, which touches on financing.

There were some occurrences where the test items and answers that could be more clarified, although they do not alter the correct answer. Some examples are presented. One item asks about the benefits of multiple sourcing. One of the correct options is "Larger economies of scale". Here, it should be clarified that it is from the supplier's side and thus not that our company gets economies of scale. In another question, the purpose is to understand that it may be a good idea to consolidate order volumes for benefits. One of the distractor alternatives is to investigate "If you can have the same supplier for all geographies exclusively ". The fault in this alternative is that having only one company does not necessarily need to be right, whilst reducing the number of suppliers is correct. This distinction should be more clearly. A final example is where the questions state certain abbreviations that should be written out. Some amounts with MUSD, i.e. Million USD, and some questions contain some financial key figures such as EBIT.

Key check

In one case, there were multiple alternatives that could be perceived as correct. The question asks one to find expected payment terms based on a supplier's financial relevant figures. The correct answer is Accounts receivable / Revenues * 365. The incorrect answer was selected as the correct one. This question also have be too difficult and unnecessary according to the PhD.

Distractor evaluation

In some cases, there were multiple alternatives that could be perceived as correct, and some examples are presented. One question asks one to choose appropriate action if one has a new supplier who might be able to replace the old one. The answer is to test the new supplier with a limited volume. Here, negotiating the price could be reasonable if there are high switching costs. One question asks how to avoid the risk of escalating costs in a contract in connection with unexpected increased material costs for the supplier. The correct answer is "To include a clause for price review in your contract with an agreed price cap that you can manage ". the purchasing PhD argues that "To set up a governance structure to allow a structured review of price change" may also be correct.

One question examines which one of several contract terms is most important to examine when evaluates a provider of a particular type of IT solution. Here is the correct answer "The handling of personal and company data". The alternative to have proper handover and code walkthrough might also be right. Proper tutorial and walkthrough is one of the pillars of handover. A final example is one question on how to act if a quotation is significantly lower than internally estimated. The answer is to get in touch and hear if the calculations are correct. Here it may also be right to review our internal assumptions.

4.3 Pilot study and item response analysis

During the pilot study, results were gathered from all 143 participants. The data will assess the suitability of different tests. To begin with, it was examined whether the test scores aligned overall with the actual competency levels of the purchasing professionals. For this, the test scores were compared with hypotheses of the purchasing managers. 8 out of 13 purchasing companies returned evaluations and the resulting evaluations compared to the scores is presented in figure 4.1, together with a linear regression.

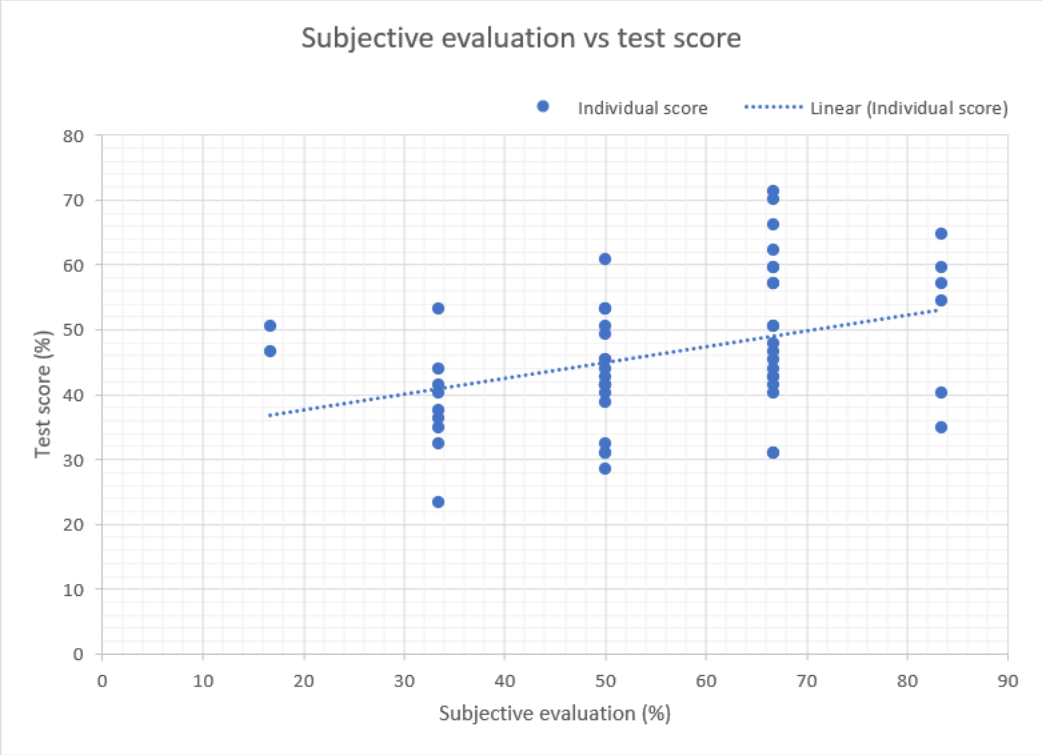


Figure 4.1: Subjective evaluation vs test score.

The resulting Pearson's correlation, the number of participants and the calculated p-value derived form the t-distribition of the Pearson's correlation is shown in table 4.1.

Table 4.1: Resulting correlation for each company and total

Company ID	Number of participants	Correlation	p-value
1	5	0,85	0.068
2	2	N/A	N/A
3	9	0,73	0.026*
4	3	0,44	0,70
5	5	0,90	0,037*
6	5	0,02	0,97
7	2	1,00	N/A
8	29	0,21	0,27
Total	**60**	**0,39**	**0,0021****

The overall correlation is positive with a significance of p=0,0021, as can be seen in the last row of the results. There is also a positive correlation for all surveyed companies individually, and with a simple binominal distribution analysis, this shows a significance of p=0,0078. The insignificant individual results are mainly due to the small sample sizes leading to high variances. This will be further discussed in the analysis chapter.

The distribution for the sujective evaluation is shown in figure 4.2 and more in-depth in table 4.2.

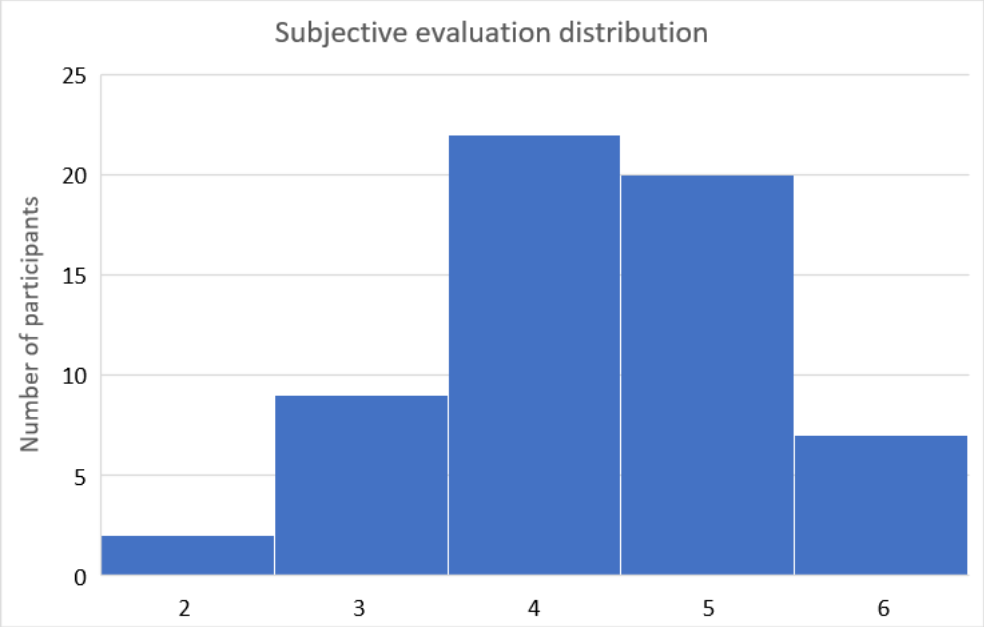


Figure 4.2: Subjective evaluation distribution.

Table 4.2: Distribution of subjective evaluations for each company.

Company ID	Number of participants for each evaluation score							
	1	2	3	4	5	6	7	Average
1	0	0	0	3	2	0	0	4,4
2	0	0	0	0	2	0	0	5,0
3	0	1	1	2	5	0	0	4,2
4	0	1	0	1	0	1	0	4,0
5	0	0	0	0	3	2	0	5,4
6	0	0	0	3	2	0	0	4,4
7	0	0	0	1	0	1	0	5,0
8	0	0	8	12	6	3	0	4,1
Total	0	2	9	22	20	7	0	4,4

The subjective evaluation scores is not spanning the entire likert scale, and the evaluations of 1 and 7 are not assigned to any of the participants. The average subjective score for each company is for the most part somehow skewed towards greater than the Likert scale average definition of 4.

4.3.1 Test score and company overall profitability

As previously stated in the theory chapter, studies have shown that purchasing competency positively affects companies' financial results. It was therefore interesting to investigate the connection between the overall test results and the corresponding financial results for the company. This connection is presented in figure 4.3.

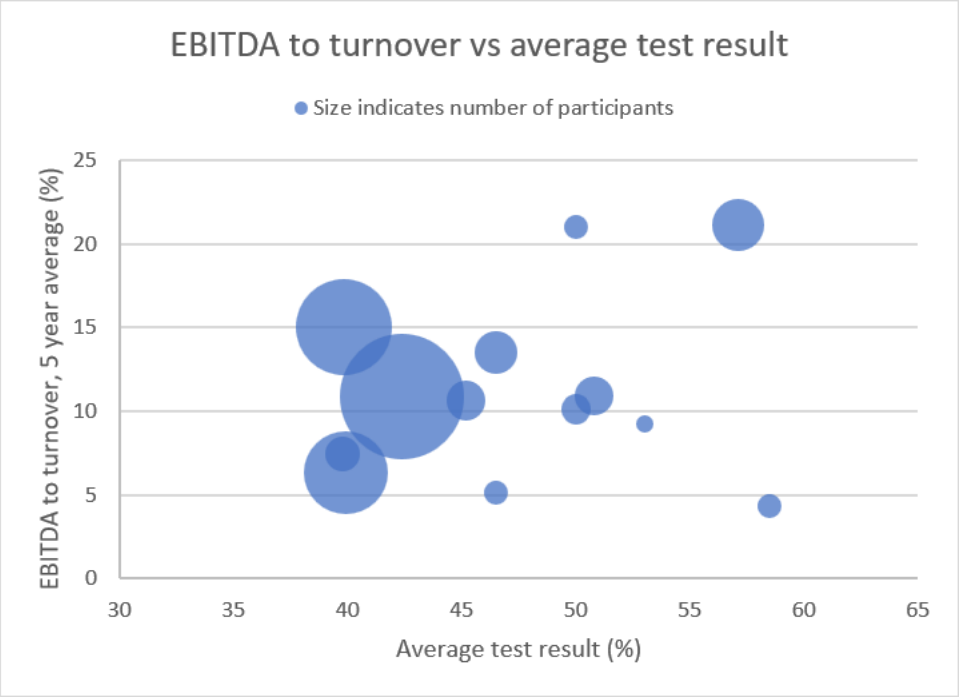
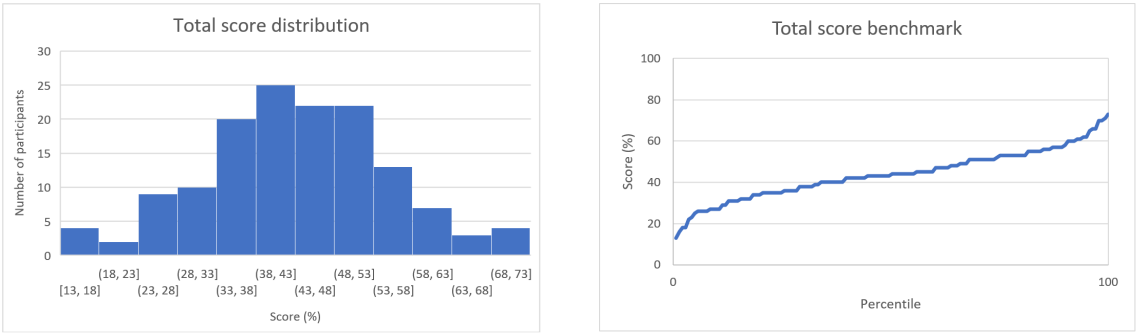


Figure 4.3: EBITDA to turnover (%) vs average test result(%).

The overall results show a positive correlation coefficient of 0,18. The significance of the correlation coefficient is tested from the p-value from the normalized t-distribution with 11 (i.e. $n-2$) degrees of freedom. The significance test shows a p-value of 0,31 which shows that the correlation is not significant. With the small sample of 13 participating companies, the variance is large enough so that this test should not be expected to produce a significant result which will be further discussed in the analysis section.

4.3.2 Resulting benchmarks

The empirical density and cumulative distribution of total score for the test is presented in figure 4.4. For the rest of the categories, the cumulative empirical distribution is presented in figure B.1 in Appendix B.



(a) Total score distribution *(b) Total score benchmark*

Figure 4.4: Empirical distribution for total score.

4.3.3 Discrimination analysis

The discrimination analysis was conducted in three ways. Firstly, each test item was compared to the rest of the items in the entire test (1 item to 76 items). Secondly, each test item was compared to the rest of the items in the corresponding area (1 item to 10 items). Finally, each test area were compared to the rest of the areas in the entire test (1 item to 6 items).

Table 4.3 shows the item-to-total correlation for each item compared to rest of items in entire test. This test tells how each item compares to the rest of the items in the entire test. A positive correlation means that the item is a contributing part in the overall assessment of the entire test.

Table 4.3: Item-to-total correlation for each item compared to rest of items in the test.

	Item-to-total correlation										
Area	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
SS	0,21	0,34	0,16	0,11	0,16	0,16	0,00	0,07	0,06	0,07	0,17
RM	0,11	0,03	0,18	0,08	0,20	0,22	0,04	-0,07	0,03	-0,01	0,16
Neg	-0,02	0,03	0,11	0,16	-0,02	0,15	-0,02	0,14	0,03	0,02	0,05
CVM	0,20	0,10	0,33	0,17	0,26	0,19	0,10	0,10	-0,03	0,09	0,28
Fin	0,20	0,24	-0,06	0,11	0,02	0,12	-0,01	0,12	0,30	0,04	0,12
Sus	0,11	0,25	0,21	0,21	0,23	-0,04	0,07	0,16	0,22	0,34	0,14
CCM	0,30	0,06	0,40	0,24	0,35	0,31	0,34	0,18	0,16	0,38	0,25

Table 4.4 shows the item-to-total correlation for each item compared to rest of items in the particular area. This test tells how each item compares to the rest of the items in the area. A positive correlation means that the item is a contributing part in the overall assessment of the area.

Table 4.4: Item-to-total correlation for each item compared to rest of items in the particular area.

	Item-to-area correlation										
Area	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
SS	0,33	0,35	0,27	0,16	0,19	0,30	0,06	0,18	0,15	0,15	0,27
RM	0,14	0,01	0,20	0,11	0,31	0,37	0,18	0,16	0,19	0,06	0,33
Neg	-0,18	-0,04	0,37	0,34	0,14	0,14	0,14	0,15	0,22	-0,11	0,12
CVM	0,16	0,20	0,34	0,27	0,34	0,29	0,34	0,10	-0,03	0,20	0,40
Fin	0,37	0,26	0,16	0,16	0,17	0,14	0,13	0,16	0,27	0,10	0,06
Sus	0,16	0,38	0,22	0,24	0,30	0,11	0,13	0,24	0,25	0,40	0,30
CCM	0,30	-0,04	0,38	0,44	0,37	0,38	0,27	0,25	0,23	0,43	0,26

Table 4.5 shows the area-to-total correlation. This test tells how each area compares to the rest of the areas in the test. A positive correlation means that the area score is a contributing part in the overall assessment of the test.

Table 4.5: Area-to-total correlation.

Area	area-to-total correlation
SS	0,60
RM	0,58
Neg	0,31
CVM	0,60
Fin	0,50
Sus	0,63
CCM	0,61

4.3.4 Item difficulty analysis

The item difficulty is measured by the portion of people manages to score correctly on the test item. The item difficulty distribution is shown in table 4.6. The difficulty distribution is also presented visually for each area in figure C.1 in Appendix C.

Table 4.6: Item difficulty. The table shows how many people managed to score correctly on the different test items.

	Item difficulty										
Area	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
SS	0,48	0,52	0,55	0,45	0,35	0,45	0,52	0,45	0,42	0,59	0,47
RM	0,26	0,20	0,50	0,11	0,33	0,41	0,75	0,43	0,47	0,31	0,50
Neg	0,55	0,46	0,45	0,50	0,97	0,40	0,84	0,18	0,23	0,38	0,04
CVM	0,39	0,79	0,33	0,75	0,10	0,38	0,45	0,26	0,26	0,28	0,41
Fin	0,72	0,35	0,28	0,11	0,25	0,34	0,31	0,43	0,29	0,74	0,34
Sus	0,56	0,25	0,54	0,32	0,69	0,40	0,08	0,26	0,65	0,57	0,25
CCM	0,52	0,55	0,79	0,51	0,52	0,88	0,75	0,44	0,45	0,55	0,21

Distractor evaluation

An analysis was also performed looking at the response distribution for each question. As stated in the theory chapter, a question that has a distractor that is chosen more often than the key might be ambiguously stated or have several correct answers. In total, 22 out of the 77 questions had a distractor that was more common than the key. Interestingly, all questions except one (Finance 7) with a difficulty of 33 percent correct or less had at least one distractor that was chosen more frequently, and none with 33 percent or greater correct answers.

4.3.5 Internal consistency analysis

Finally, the internal consistency was calculated with the KR-20 measure for binary variables, and is shown in figure 4.7.

Table 4.7: KR-20 measure for each tested area as well as for the entire test.

Area	KR-20 measure
SS	0,37
RM	0,27
Neg	0,18
CV	0,43
Fin	0,35
Sus	0,45
CCM	0,59
Total	**0,82**

The internal consistency shows KR-20 values for the individual areas ranging from 0,18 to 0,59. Based on the requirements of 0,7 for a reliable test according to the rule of thumb. The total score, with 77 test items, show a value of 0,82 which is greater than the required value of 0,8 and thus it is shown to be reliable according to the rule of thumb.

5

Analysis

The results of the content review and pilot study gave an insight in how successful the test is in determining the validity of the MCQ test for the chosen competency areas and test test cases, and the analysis is presented below.

5.1 Analysis and conclusion of relevant competency areas

The relevant competency areas hypothesized in the theory chapter were investigated by cross-case analysis of purchasing course contents and content reviews with three experts in the field of purchasing.

5.1.1 Cross-case analysis of course curricula

The competency areas that were hypothesized each has brief descriptions presented in the theory chapter. Since the course curricula contain rather broad descriptions of areas covered by the course, the contents of the areas was more specifically analyzed in the content reviews. The purpose of analyzing the curricula was foremost to assess if the topics were present in a large portion of the courses and therefore would indicate that these areas are important for strategic purchasing.

The course curriculum investigation showed seven focus areas that were present in a majority of the courses. These areas were overall understanding of the purchasing role, purchasing strategies, supplier relationship management, sustainability, negotiation, cost management, and public procurement. The deductive approach of assessing the chosen competency areas made it relevant to go into each of the hypothesized areas and evaluating the presence of each one.

5.1.2 Content reviews

In the content reviews and validity analyses, a more complete analysis of the framing of content of the competency categories was conducted. First of all, **Sourcing strategy** is mentioned in a majority of the courses, and the different sourcing strategy related contents of the course curricula is overall in line with with the sourcing strategy contents in the theory chapter. Thus, it can be validated as being a relevant area. **Negotiation** is mentioned in a majority of the courses and which indicates that it is a relevant area. The contents of negotiation preparation and negotiation per se is in line with the contents of the area. **Contracting & Contract management** is only explicitly mentioned in one of the courses. However, the activity of supplier relationship management was mentioned in all five course curricula. This activity is overlapping with Contracting & Contract management to a quite large extent. It could be argued that for example supplier improvement falls under supplier relationship management as well, but in our case it falls under sourcing strategy due to the value of supplier improvement possibilities in strategic sourcing.

Sustainability is shown to be relevant as it was present in the majority of courses. Similarly to the theory chapter, there are several different notions for the subject of sustainability. This further justifies the notion of simply using sustainability since any specific notion such as PSR, SRP, EPSM or SEEB would be unfamiliar to many. **Cost & Value management** is explicitly mentioned in a majority of course curricula. This is not surprising as cost management as a main objective of the purchasing function. The relevance of the value management part of the competency would be further investigated in the content reviews, but since value analysis and value engineering is an established subject in purchasing according to the theory chapter, it is not rejected in this stage. As a consequence however, going into the design phase of the competency test, a larger portion of the test cases was focused towards cost management in order to give a reflection of real life application of competencies. **Risk management** is only explicitly mentioned in two of the five course curricula. However, due to its intrinsic value in all parts of the purchasing process, it is implicitly mentioned in several course curricula, for example as critical evaluation of purchasing practices and situations.

financing is only included in one out of the five curricula. The relevance of the competency is therefore not verified. By studying the prerequisites for the purchasing courses, which is highly related to the relevant competencies, one can see that financing was somehow represented through Industrial Economics. However, this area is not verified strongly as a relevant area in this analysis and it was therefore a focus to investigate in the content reviews.

In addition to the hypothesized competency areas, the only highly present areas that are not covered by the hypothesized areas is **public procurement** and **understanding the role of purchasing**. It can be further investigated if these are also relevant competency areas for assessing the strategic and analytic skills and abilities to make correct judgments based on data in a purchasing setting. This is focusing on finding some relevant areas, and the question whether these are the most relevant areas of all is a relevant subject for further studies.

To summarize, all hypothesized competency areas are present more or less in the course curricula, although financing and value management is less present than the others. Based on the course curriculum analysis, the areas that are less mentioned might be considered to either be weighted less in the overall judgement, or being extended to covering more similar areas. However, since none of the areas are clearly redundant and all areas covered distinct objectives, activities and qualifications, these were not altered and also weighted equally going forward.

5.1.3 Content reviews on the chosen competency areas

In the content reviews, all three participants are in unison that the seven competency areas as stated in their description are relevant "strategic and analytic skills and abilities to make correct judgments based on data in a purchasing setting" for strategic purchasing professionals. In addition, there are no important and significantly large areas that are not covered according to the experts.

It was highly relevant to discuss with the purchasing PhD why financing and risk management aspects of purchasing not being as present in the courses, since the individual is administering the purchasing course at LTH. It is confirmed in the content review that these are actually two important aspects within purchasing, and that students had previously specifically asked to include more of these aspect in the purchasing courses.

5.1.4 Final choice of competency areas

The analysis of course curricula and the comments from the experts of the proposed competency areas are all in line with the hypotheses of the seven competency areas. indicating a positive result of the content validity of both test areas and test cases. This confirms the first sub-hypotheses that each competency area is relevant for strategic purchasers (H1a, H2a, ..., H7a).

5.2 Analysis and final choice of MCQ test test cases

With the first hypothesis confirmed for each individual test area, the next hypothesis relates to the remaining validity test structured in figure 3.4.

5.2.1 Validity and final choice of test areas

To begin with, the overall positive item-to-total correlation between the test item and the total score, as well as the overall positive area-to-total correlation, foremost shows that participants tend to score similarly well on all areas. As all areas assess ability to make the strategic and analytic decisions based on data, it

can be expected that some participants are more analytically minded overall, and thus perform more well in tests overall. This also raises the hypothesis that purchasing professionals with more analytic mindset overall make more analytically favorable and pragmatic decisions in practice.

There is still a distinction between the categories with correlations ranging from 0,31 to 0,61. According to Considine et al. (2005), if a correlation between an item and the total score is greater than 0,70, the item risks being redundant as it does not distinct between the other test cases enough. Treating one category as a test item, this shows that there is a distinction between the categories and that the participants did perform differently in the different competency areas. The groups of questions in each test area are therefore distinct enough so that none of the test areas should be rejected as redundant.

5.2.2 Validity and final choice of MCQ test cases

Qualitative requirements

First of all, since the content reviews were performed through several discussions and re-iterations, all comments on content validity, clarity, keys and distractors could be re-iterated and resolved. Due to this, all final test cases passed the content validity, face validity, key check, and distractor evaluation.

Distractor evaluation

The more difficult questions most often have at least one alternative that was considered more plausble than the rest of the incorrect alternatives. Looking at the incorrect questions, there are three main findings.

The first finding is simply that one of the alternatives was found a more popular pick. For example, in one sustainability question stating how to make sure a supplier agrees to your company's environmental and social demands, making sure that the company was iso14001 certified was more popular than assessing if the supplier agreed to our code of conduct. This example shows that some incorrect alternatives might sound correct and thus distinct the knowledgeable, which is a trait for a good distractor according to Considine et al. (2005). The second finding was that most of these questions were simply difficult or somewhat ambiguous. Several questions required calculations and the difficulty indicates that the questions were too difficult for many in the given time frame.

The 22 difficult questions were discussed once more with one of the experts to assess if any of the questions had other correct answers. Although some formulations were a tad bit ambiguously stated in hindsight, they were still concluded to have one unique correct answer. For these areas, the discrimination index would therefore sufficiently determine whether the question managed to distinct the more knowledgeable within the field to the less.

Discrimination analysis

As stated in the methodology chapter, the final choice of MCQ test cases should be selected based on the discrimination indices (Considine et al., 2005). A high discrimination index means that the test item can discriminate between the participants that perform well overall in the area to the participants scoring less well. Since the test consists of seven distinct areas, there are no expectations on the test that participants scoring well on one category necessarily needs to score well in other categories. The seven test competency areas should therefore be viewed as seven distinct MCQ tests. The test cases were therefore analyzed based on how well they distinct the participants in the certain competency area, i.e. the item-to-area correlation.

Overall, 72 out of 77 questions show a positive correlation and contributes to the overall assessment. According to Considine et al. (2005), an item-to-total correlation of 0,25 or greater is considered acceptable. This measurement is quite harsh and would render 31 out of 77 questions valid. With the sample of 143 participants, the estimated correlation coefficients are prone to uncertainty depending on sample size, and simple two-sided confidence interval calculation for the Pearson correlation coefficient shows that test cases with a correlation of 0,09 has a 95 percent confidence interval of [-0,07526; 0,2504] and thus cannot be rejected with significance 0,05. Therefore, the 68 questions with correlation $r \geq 0,09$ cannot be rejected and a fair deduction is that they could be kept until studies with additional participants have been done.

As stated in the methodology chapter, the remaining 9 test test cases should be eliminated or replaced with ones with a desirable difficulty. As is seen when comparing the difficult questions discussed in section 4.3.4 to the item-to-area correlations in 4.4, several of the most difficult questions are the ones with a low item-to-area correlation, and the remaining difficult questions therefore successfully distinguish between competencies in the higher competency spectrum. Based on the results in table 4.6, the difficulty is generally biased towards the range of 50 percent or less correct answers. This means that in general, test cases in the simpler spectrum with expectation of greater than 50 percent correct answers should be included to decrease this bias and allow for better discrimination in the lower competency spectrum.

5.3 Validation of the test purpose

5.3.1 Overall test validity

The comparison between the subjective evaluation and test score presented in table 4.1 shows a positive correlation with a strong significance of p=0,0021. This rejects the null hypothesis for H8a. The next finding is that the individual correlation shows exclusively positive values for each of the seven individual companies with measurable correlation, yielding a significance of p=0,0078. This means that also the null hypothesis for H8b is rejected. This confirms with strong

significance the hypothesis H8 and the purpose of successfully measuring the strategic and analytic skills and abilities to make correct judgments based on data in a purchasing setting.

Both the evaluation and total score follow a similar distribution, resembling the normal distribution. This does not mean any further connection between the two in itself, but it does reduce the variance of the the correlation between the two and thus further increases the validity of the explanation value of the correlation coefficient (Pharr et al., 2021).

5.3.2 Test validity for each area

First of all, all test areas show an acceptable distinction, and the test itself shows to fit its purpose with a strong significance. All qualitative validity requirements are passed indicating further that the test test cases assess each individual competency area. With the rejected questions removed due to significantly weak discrimination, a remaining 68 questions pass all the validity requirements to the extent that they cannot be rejected. This remaining test pruned down to 68 questions therefore confirms the second sub-hypotheses the areas will successfully be measured by the test (H1b, H2b, ..., H7b). Finally, the seven hypotheses that each test category is valid (H1, H2, ..., H7) is accepted for all seven areas.

5.4 Reliability of the test

With the confirmed validity that the test assesses competency overall and for each individual area, the reliability of the test is highly relevant to assess. The internal consistency for the test in total of KR-20 = 0,82 indicates that the test is suitable for assessing the overall competencies in a reliable way, above the recommended value of 0,8. For each individual area, the internal consistency shows values ranging between 0,18 and 0,59 and the test is therefore not suitable for assessing individual scores for individual areas in a fully reliable way. It is important to mention again that the reliability measurement requirement is a rule of thumb, and the test does have applications anyways, which will be discussed in section 7.3.

Firstly, a test with high margin of error still shows a reliable result within a certain granularity. For example, a score in the 0-20 percentile can still be shown to reliably distinguish the participant from the ones with a score in the 80-100 percentile. Secondly, the test can still be shown to be reliable for a group of participants. The results for a group of professionals is more reliable than for a certain individual, as the results converge to a mean with an increasing sample size. The purpose of assessing a purchasing function in a company consisting of a group of purchasers is therefore increasingly fit for its purpose given that the number of participants is high enough. More investigation will not be, but as the variance decreases proportionally with sample size, it is not surprising if a group of three or more participants would lead to a reliable overall measurement in each individual area.

5.5 Test result and company overall profitability

The correlation between the test score and financial results is positive, but not significant. Any significance is not expected to be found due to the small sample of 13 companies participating. There is also an intrinsic variance in each of the participating companies that should be taken in mind. Firstly, companies with a larger number of participants should be weighted higher, and secondly the internal variance of individual company score should be taken in mind. The profitability also varies significantly between industries, and therefore a more rigorous analysis would require a larger sample of companies from each industry.

This result is therefore not of any significance as of this sample size, and only presents as a proof of concept that this correlation is in line with the previous studies. A more rigorous analysis and significant results would require more participating companies.

6

Discussion

6.1 Determining relevant competency areas for strategic purchasers

To begin with, the hypotheses appeared to be at the same level of depth as the elements offered in the course curricula, enabling an analysis with clear comparisons and little need of personal assumptions. However, there were some assumptions made on the groupings of certain topics. Secondly, the course curricula were in fairly good agreement with one another and with the hypothesized areas. This suggests that the chosen competency areas are not overly specific or narrow, as the content reviews verified.

It needs to be mentioned that the detail level varied between the different course curricula. It can be assumed that more of the courses involve the different elements, and also that there might be additional important areas covered that cannot be traced to any of the chosen competence areas. A more detailed investigation here would be to combine the course curriculum analysis and interviews with the corresponding course coordinator.

The courses were not delimited to strategic purchasing exclusively, for example operative purchasing and more overall business strategies. It is therefore not entirely certain that all the curricula contents are present in a strategic purchasing context. However, the course descriptions and discussions with the purchasing PhD administering one of the courses both suggest that all of the strategic purchasing competencies are present in the courses.

The chosen test scenarios fairly comprehensively analyzed the competency areas with a fair weight based on the interviews and course curriculum analysis. However, the content assessments were mostly undertaken with three specialists, with additional brief discussions with other experts. Three extensive content reviews were the minimum according to Considine et al. (2005), and by discussing with more experts, there might arise other opinions of what requirements should be weighed in.

6.2 Designing a test to measure competencies

6.2.1 Finding purchasing professionals to participate in the pilot study

The sampling method used for finding purchasing professionals leads to some bias when not taking all company sizes and geographies into consideration. There were mainly Swedish mid- and large cap companies participating, making the assessment more uncertain towards smaller companies. For example, smaller companies might have other priorities or expertise and there is a lack of benchmark between these sizes. An interesting next step could be to allow for benchmark between similar company sizes and industries, but that is outside of the scope of the purpose of this study is to assess cross-industries and sizes.

The purchasing professionals making up the pilot study was not solely strategic purchasing professionals. This was resolved by asking the participants to specify What best matched their job title, and only input several if each is ≥ 30 percent of their time. It was shown that nine out of the 143 purchasing professionals are operational purchasers. In the analysis, these were included but in the final benchmark, these should be excluded since they are not included in the scope of the benchmark study.

6.2.2 Interdependence between participants

An important aspect to take in mind throughout the final benchmark and the overall analysis of the validity and reliability is the interdependence between participating purchasing professionals. In this study, each individual participant is assumed to be independent of any other participant. The 143 participants are however grouped into 13 companies, and it can be assumed that there are some commonalities between these companies such as similar activities, knowledge sharing between professionals, and on-the-job training efforts of the particular company. This leads to the results possibly being somewhat correlated within the company. It could also be the case that individual companies are overall stronger or weaker in some of the purchasing competency areas.

Since a large portion of participants comes from three companies, the risk of this kind of bias is rather large. For a more unbiased benchmark and analysis, the results could be aggregated into 13 companies. For this grouping to produce significant results, the a focus for future expansion of the study would be to find more participating companies. Right now, a consideration to use the aggregated results from the 13 companies as benchmark is still valid.

6.2.3 Subjective evaluation

The fact that the average correlation for each single company is greater than the overall correlation indicates that some of the variance that is not explained by the

test score might be explained by the incoherence in the expected average performance of each single company. It can be discussed whether the evaluated competencies should be normalized, or if the purchasing managers should have received additional directives. However, the brief instructions allows for the most objective assessment without outside intervention.

Looking at the subjective evaluations in figure 4.2, it is evident that the subjective evaluation follows a normal distribution. This in itself is an interesting finding, suggesting a normally distributed perceived competence in the purchasing workplace. However, it is also evident that no managers set either a lowest of highest value on any of the purchasers, which would provide a more detailed explanation. There are several ways this could have been resolved beforehand, for example use less distinct wordings than ”1. Poor” and ”7. excellent”. These wordings are also not entirely in line with a seven point likert scale, since the extreme points are not fully symmetric which could affect both the choice of scoring and the symmetry.

6.2.4 Final choice of MCQ areas and items

There are several points of view on what test items should be rejected until further investigation. First off, in this case the questions who's discrimination cannot be rejected as being below 0.25 are kept and the remaining rejected. If no further analysis will be made, it can also be argued that only the items that discriminate over 0.25 with a significance which would lead very few question but still a highly valid test. Finally, the value of 0.25 is also quite harsh and it can be argued that as long as the item has a significantly positive correlation, it does provide to the discrimination and should be kept.

As discussed in the content validity, a few questions might have been a better fit in other categories. This alteration can actually be performed after the pilot study and validity test results reevaluated. For the final choice of mcq-items, it could therefore also be relevant to investigate whether some of the questions with a low discrimination value would be a better fit with higher discrimination in any of the other categories, before simply eliminating the test item. These alterations should however only be performed if it is qualitatively assumed, since alterations simply based on the discrimination is prone to p-hacking.

6.2.5 Validity of the tool

For the entire test, both qualitative and quantitative measures were made to validate the tool's fitness for purpose. The same goes for each test item, where both qualitative and quantitative validity measures were taken. The ability for each test area to measure a single competency is however only validated qualitatively through content reviews. One interesting analysis would be to also perform a construct validity measurement for each test area. This would be performed by asking purchasing managers to subjectively evaluate the perceived competency of each participant in each specific area and analyzing the correlation

between the evaluations and test results. However, the subjective evaluations would probably be hard to evaluate by the managers due to difficulty in both framing each area as well as evaluating each sub-competency of each strategic purchaser. Therefore, the validity test would not be meaningful with this sample size to test a hypothesis.

7

Conclusion

The chapter will revisit the purpose, which is to create and validate a tool that measures the extent to which strategic purchasers have some competencies relevant to high performing purchasing and supply chain organizations.

7.1 How can a suitable number of competency areas that are relevant for the performance of strategic purchasing professionals be identified?

To conclude, seven relevant competency areas have been found through literature review of activities and objectives of strategic purchasers the competency of best practice books for strategic purchasers as well as existing studies on the overall competency is important for strategic purchasers. These have later been validated through cross-case analysis between the course curricula of five Swedish advanced university courses in purchasing, as well as interviews with three experts within the field of strategic purchasing. Similarly, the content of each area has been found through literature review and validated through both qualitative measures and quantitative measures in the form of content review with three experts and analysis of the distinction between the areas.

Using this process, the seven areas that were identified was sourcing strategy, negotiation, contracting contract management, cost value management, risk management, sustainability, and financing. The contents of each of these areas are defined in the theory chapter and a conceptual framework for the interrelation between the areas is proposed and used for determining relevant requirements for assessments is found in figure 3.3.

7.2 How can a tool be created for an objective and structured assessment of these competency areas?

For creating a tool that accurately assess the proposed competency areas and their contents, a requirement specification was set up with a mix of requirements based on the contents of the area. One corresponding test case was set up for each requirement. The mix of test cases for each area went through several iterations of content review with experts in order to be considered suitable for comprehensively and fairly assessing each proposed competency area.

After several iterations, each test case was considered relevant for strategic purchasers and the mix of test cases fair for assessment of each area. Thus the qualitative validity requirements for the test itself tested areas respectively as well as each test had been fulfilled, suggesting that the test does evaluate each unique competency area. Whether the test cases were clear and specific enough to be able to distinguish between competency levels in practice the test still needed to be practically evaluated through a pilot study.

7.3 How can the validity of the test be extensively assessed?

In order for the test to be valid, the test itself as well as each test case needed to be evaluated through several validity measures. The results of the validity analyses suggests that the test itself proves to be valid with strong significance on both hypotheses. All test areas also demonstrated appropriate differentiation.

For the test items, 9 test items should be rejected with the sample size used since the item-to-area correlation was significantly below the required value of 0,25. The 68 questions with an item-to-area correlation of 0,09 or more are shown valid to the extent that no questions can be rejected as not discriminatory enough with significance, and should be kept until further studies have been made. For further unbiased validity and reliability, the tested data points should be aggregated into the average performance for each company.

With the test shown to be valid, the reliability of the test was analyzed through calculating the internal consistency. From this analysis it can be concluded that the tool has these four characteristics for reliability:

- The tool is reliable for assessing overall competencies for a group.
- The tool is reliable for assessing overall competencies for an individual.
- The tool is reliable for assessing competencies within a specific area for a large enough group. Further investigation on the group size is needed.

- The tool is not reliable for assessing competencies within a specific area for an individual. Further investigation if the test can reliably distinguish top and bottom performers is needed.

7.4 Practical Implications

This test has been shown to successfully analyze the competencies of strategic purchasers. The four characteristics presented in the end of the last section gives a description of the the practical value of the test.

The discussions on expectations, results, and possibilities during the results presentations that were held for the 14 companies showed several practical uses of this data-driven and objective measurement. Discussions were held whether the test could be used for recruitment purpose, but foremost it enables a pre-study that is useful for competency improvement efforts of individuals, purchasing offices, and the purchasing organization in the company overall.

The ability to assess the how competencies varied within the company, and also compared to other companies, was especially valueable for the participating companies. For purchasing organizations with few strategic purchasers, it was especially important to be able to objectively assess competency compared to a larger benchmark.

7.5 Theoretical Implications

The purchasing assessment tool is mainly following the design and validity process of Considine et al. (2005). This process was focused on nursing and education, but it has been shown that this process has been shown to be applicable for strategic purchasing and, in a sense, also education. For puchasing, the tool has contributed to the the development of the field of knowledge of purchasing performance management.

The requirement specification has been shown to be suitable for quality assurance of successful strategic purchasing and education in a sense. This thesis contributes to the usefulness of the structure of requirement specifications for quality assurance for education and best practice work assessment.

7.6 Future Research

For this particular tool, it can be further investigated if any of the additional competence areas that were common in the course curriculum analysis could be added to further increase the diversity of the tool. The weight of each competence area for overall evaluation can also be further investigated to give a more fair assessment of overall competence.

This was a Cross-sectional study, leaving out some of the longitudinal reliability measurements. It would therefore be interesting for future research to complete the entire study by incorporating the longitudinal reliability measurements.

In addition to financial performance, there are other interesting validity measurements that would require a larger sample of companies or participants in order to be meaningful. One is the relationship between assessment performance and company profitability, reduction in costs of goods sold, resilience towards economic downturns, or other relationships proposed by previous studies. Another is to compare the score with the perceived competency in a certain area to further validate the test purpose.

This thesis also contributes to the practical use of objective assessment tools for best practice work and pre-studies for competency improvement. It is highly plausible that future tools can designed in the same way for valid, reliable, and useful assessments of other business functions.

References

AE University. What is difference between finance and financing, Accessed 2022. URL https://www.svtuition.org/2011/08/what-is-difference-between-finance-and.html. Accessed 2022/05/21.

American Society for Quality. Quality assurance quality control, 2022. URL https://asq.org/quality-resources/quality-assurance-vs-control. Accessed 2022/05/21.

P. Anselmi, D. Colledani, and E. Robusto. A comparison of classical and modern measures of internal consistency. *Frontiers in Psychology*, 10, 2019. URL https://www.frontiersin.org/article/10.3389/fpsyg.2019.02714.

D. Aulia and W. Isvara. Strategies to increase procurement maturity level using procurement maturity model to improve procurement performance. *International Journal of Scientific and Research Publications (IJSRP)*, 11:489, 06 2021.

J. Backstrand, R. Suurmond, E. Van Raaij, and C. Chen. Purchasing process models: Inspiration for teaching purchasing and supply management. *Journal of Purchasing and Supply Management*, pages 100–577, 10 2019.

C. Baier, E. Hartmann, and R. Moser. Strategic alignment and purchasing efficacy: an exploratory analysis of their impact on financial performance. *Journal of Supply Chain Management*, 44(4):36–52, 2008.

L. Bals, H. Schulze, S. Kelly, and K. Stek. Purchasing and supply management (psm) competencies: Current and future requirements. *Journal of Purchasing and Supply Management*, 25(5):100572, 2019. ISSN 1478-4092.

P. Barrett. Kr20 coefficient alpha: Their equivalence for binary-scored items. 2007. URL https://www.pbarrett.net/techpapers/kr20.pdf.

C. Bayazit, P. Stanis, M. Lauritzen, and B.-U. Mercker. Assessing individual capabilities in the procurement function. Web page, 2015. URL https://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/mckinsey%20capability%20building/pdfs/mcba%20-%20procurement%20brochure.ashx. Accessed 2022/05/21.

M. Borg, J. Bengtsson, H. Österling, A. Hagelborn, I. Gagner, and P. Tomaszewski. Quality assurance of generative dialog models in an evolving conversational agent used for swedish language practice, 2022. URL https://arxiv.org/abs/2203.15414.

A.-M. Brady. Assessment of learning with multiple-choice questions. *Nurse Education in Practice*, 5(4):238–242, 2005. ISSN 1471-5953. doi: https://doi.org/10.1016/j.nepr.2004.12.005. URL https://www.sciencedirect.com/science/article/pii/S1471595305000065.

S. Buckles and J. J. Siegfried. Using multiple-choice questions to evaluate in-depth learning of economics. *The Journal of Economic Education*, 37(1):48–57, 2006.

F. Caniato, D. Luzzini, and S. Ronchi. Purchasing performance management systems: an empirical investigation. *Production Planning & Control*, 25(7): 616–635, 2014.

A. Carr and L. Smeltzer. An empirical study of the relationships among purchasing skills and strategic purchasing, financial performance, and supplier responsiveness. *Journal of Supply Chain Management*, 36:40–54, 06 2000.

A. S. Carr and L. R. Smeltzer. The relationship of strategic purchasing to supply chain management. *European Journal of Purchasing Supply Management*, 5(1): 43–51, 1999. ISSN 0969-7012.

C. Carter. Purchasing social responsibility and firm performance: The key mediating roles of organizational learning and supplier performance. 2005.

C. R. Carter and M. M. Jennings. The role of purchasing in corporate social responsibility: A structural equation analysis. *Journal of Business Logistics*, 25 (1):145–186, 2004.

CIPS. Socially responsible procurement - corporate social responsibility in a supply chain context, Accessed 2022. URL https://www.cips.org/knowledge/procurement-topics-and-skills/sustainability/corporate-social-responsibility-csr/socially-responsible-procurement---corporate-social-responsibility-in-a-supply-c#:~:text=In%20broad%20terms%2C%20CSR%20is,social%2C%20environmental%20and%20business%20areas.&text=Where%20CSR%20practices%20relate%20to,performance%20in%20the%20supply%20chain. Accessed 2022/05/21.

J. Considine, M. Botti, and S. Thomas. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. *Collegian*, 12(1):19–24, 2005. ISSN 1322-7696.

P. Cousins, R. Lamming, B. Lawson, and B. Squire. *Strategic supply management: principles, theories and practice*. Pearson Education Limited, 2008.

D. K. Deardorff. *The SAGE handbook of intercultural competence*, chapter 28: Implementing intercultural competence assessment. SAGE Publications, Inc, 2009.

Department of Housing and Public Works. Supply market analysis, 2010. URL https://www.hpw.qld.gov.au/__data/assets/pdf_file/0018/3357/procurementguidesupplymarketanalysis.pdf. Accessed 2022/05/21.

DILF. Competency assessment, Accessed 2022. URL https://www.dilf.dk/en/tools/competency-assessment/. Accessed 2022/05/21.

M. Eatough. Leaders can no longer afford to downplay procurement. Web page, 2014. URL https://hbr.org/2014/03/leaders-can-no-longer-afford-to-downplay-procurement. Accessed 2022/05/21.

P.-E. Eilström and H. Kock. Competence development in the workplace: concepts, strategies and effects. *Asia Pacific Education Review*, 9(1):5–20, Feb 2008. ISSN 1876-407X.

S. C. Ellis, R. M. Henry, and J. Shockle. Buyer perceptions of supply disruption risk: A behavioral view and empirical assessment. *Journal of Operations Management*, 28(1):34–46, 2010. ISSN 0272-6963.

L. M. Ellram. Purchasing: The cornerstone of the total cost of ownership concept. *Journal of business logistics*, 14(1), 1993. ISSN 0735-3766.

L. M. Ellram and A. Carr. Strategic purchasing: A history and review of the literature. *International Journal of Purchasing and Materials Management*, 30 (1):9–19, 1994.

A. Fink. Survey research methods. In P. Peterson, E. Baker, and B. McGaw, editors, *International Encyclopedia of Education (Third Edition)*, pages 152–160. Elsevier, Oxford, third edition edition, 2010. ISBN 978-0-08-044894-7.

R. Fisher, W. Ury, and B. Patton. *Getting to Yes: Negotiating Agreement Without Giving In*. Penguin Books, New York, 3 edition, 2011. ISBN 978-0-14-311875-6.

R. N. Forthofer, E. S. Lee, and M. Hernandez. 7 - interval estimation. In R. N. Forthofer, E. S. Lee, and M. Hernandez, editors, *Biostatistics (Second Edition)*, pages 169–212. Academic Press, San Diego, second edition edition, 2007. ISBN 978-0-12-369492-8. doi: https://doi.org/10.1016/B978-0-12-369492-8.50012-1. URL https://www.sciencedirect.com/science/article/pii/B9780123694928500121.

L. Giunipero, R. Handfield, and R. Eltantawy. Supply management's evolution: Key skill sets for the supply manager of the future. *International Journal of Operations Production Management*, 26:822–844, 07 2006.

L. C. Giunipero and D. H. Pearcy. World-class purchasing skills: An empirical investigation. *Journal of Supply Chain Management*, 36(3):4–13, 2000.

GLOMACS. Strategic cost and value management in procurement, 2022. URL https://glomacs.com/training-course/strategic-cost-and-value-management-brin-the-supply-chain. Accessed 2022/05/21.

J. Hallikas and K. Lintukangas. Purchasing and supply: An investigation of risk management performance. *International Journal of Production Economics*, 171:

487–494, 2016. ISSN 0925-5273. URL https://www.sciencedirect.com/science/article/pii/S0925527315003400. Challenges for Sustainable Operations – Selected papers of ICPR 2013.

R. B. Handfield, P. D. Cousins, B. Lawson, and K. J. Petersen. How can supply management really improve performance? a knowledge-based model of alignment capabilities. *Journal of Supply Chain Management*, 51(3):3–17, 2015.

Inköpsdesign. Strategisk inköpare. Web page, 2022. URL https://www.inkopsdesign.se/inkopsroller/strategisk-inkopare/. Accessed 2022/05/21.

A. Joshi, S. Kale, S. Chandel, and D. K. Pal. Likert scale: Explored and explained. *British journal of applied science & technology*, 7(4):396, 2015.

N. L. Kendt and D. O. Nichols. *Management Responsibilities and Strategies*, chapter 14: Value Analysis and Puchasing Research. Queensland Government, Accessed 2022. URL https://www.hpw.qld.gov.au/__data/assets/pdf_file/0018/3357/procurementguidesupplymarketanalysis.pdf. Accessed 2022/05/21.

K. Kiss, C. Ruszkai, and K. Takács-György. Examination of short supply chains based on circular economy and sustainability aspects. *Resources*, 8(4), 2019. ISSN 2079-9276. URL https://www.mdpi.com/2079-9276/8/4/161.

H. Kock and P.-E. Eilström. *Competence development in the workplace: Concepts, strategies and effects*, volume 9, pages 34–54. 05 2009. ISBN 10 0-415-49210-6.

D. Lowe. *The Wiley guide to project technology, supply chain procurement management*, chapter 14: Contract Management. Wiley guides to the management of projects. John Wiley Sons, 2007. ISBN 047022682X.

D. Luzzini and S. Ronchi. Cinderella purchasing transformation: linking purchasing status to purchasing practices and business performance. *Production Planning & Control*, 27(10):787–796, 2016.

S. McCombes. Sampling methods — types and techniques explained, 2022. URL https://www.scribbr.com/methodology/sampling-methods/. Accessed 2022/05/21.

Medius. Boost your business incomes with procurement and finance collaboration, 2022. URL https://www.medius.com/blog/how-finance-and-procurement-collaboration-boosts-business-outcomes/. Accessed 2022/05/21.

B. Meger. A critical review of competency-based systems. *Human Resources Professional*, 9:22–25, 1996.

P. Paranikas, G. P. Whiteford, B. Tevelson, and D. Belz. How to negotiate with powerful suppliers. Web page, 2015. URL https://hbr.org/2015/07/how-to-negotiate-with-powerful-suppliers. Accessed 2022/05/21.

M. Pharr, W. Jakob, and G. Humphreys. Importance sampling. Web page, 2021. URL https://www.pbr-book.org/3ed-2018/Monte_Carlo_Integration/Importance_Sampling. Accessed 2017/01/25.

M. Pohl and K. Förstl. Achieving purchasing competency through purchasing performance measurement system design—a multiple-case study analysis. *Journal of Purchasing and Supply Management*, 17(4):231–245, 2011. ISSN 1478-4092.

R. Purtilo, A. Haddad, and R. Doherty. increase your ability to work with them effectively. *Career Development for Health Professionals: Success in School & on the Job*, page 198, 2015.

N. J. Salkind. Encyclopedia of research design au, 2010.

M. Saunders, P. Lewis, and A. Thornhill. *Research Methods for Business Students*. Pearson Education Limited, 5 edition, 2009.

L. Smeltzer and S. Siferd. Proactive supply management: The management of risk. *Journal of Supply Chain Management*, 34:38 – 45, 04 2006.

H. Snyder. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104:333–339, 2019. ISSN 0148-2963.

S. Talluri and R. Narasimhan. A methodology for strategic sourcing. *European Journal of Operational Research*, 154(1):236–250, 2004. ISSN 0377-2217.

F. Tangianu, A. Mazzone, F. Berti, G. Pinna, I. Bortolotti, F. Colombo, C. Nozzoli, M. Regina, A. Greco, C. Filannino, M. Silingardi, and R. Nardi. Are multiple-choice questions a good tool for the assessment of clinical competence in internal medicine? *Italian Journal of Medicine*, 12:88, 06 2018. doi: 10.4081/itjm.2018.980.

R. Tassabehji and A. Moorhouse. The changing role of procurement: Developing professional effectiveness. *Journal of Purchasing and Supply Management*, 14: 55–68, 03 2008.

W. L. Tate, L. M. Ellram, and K. J. Doole. Environmental purchasing and supplier management (epsm): Theory and practice. *Journal of Purchasing and Supply Management*, 18(3):173–188, 2012. ISSN 1478-4092.

University of Texas School of Information. Writing goals, objectives, and activities, Accessed 2022. URL https://extranet.winnefox.org/sites/extranet.winnefox.org/files/WritingGoalsObjectivesAndActivities.pdf. Accessed 2022/05/21.

A. J. Van Weele. *Purchasing and Supply Chain Management*. Cengage Learning, 6 edition, 2009.

A. Vuong, T. Nixon, and B. Towle. A method for finding prerequisites within a curriculum. In *EDM*, pages 211–216, 2011.

M. Wadi, A. F. A. Rahim, M. S. B. Yusoff, and K. A. Baharuddin. The effect of mcq vetting on students' examination performance. *Education in Medicine Journal*, 6(2), 2014.

R. B. Waina. *Specification of Educational Objectives for System Evaluation.* Santa Monica, CA, 1969.

W. B. Walstad and W. E. Becker. Achievement differences on multiple-choice and essay tests in economics. *The American Economic Review*, 84(2):193–196, 1994. ISSN 00028282. URL http://www.jstor.org/stable/2117827.

J. Webb. What is the kraljic matrix?, 2017. URL https://www.forbes.com/sites/jwebb/2017/02/28/what-is-the-kraljic-matrix/?sh=5353cb4f675f. Accessed 2022/05/21.

R. K. Yin. *Case Study Research.* SAGE Publications, Inc, 5 edition, 2009.

Appendix A

Interview guide for content review

Face validity

For each item:

- Is a realistic scenario used?
- Is the question and the alternatives clearly stated?

Content validity

General about the tested area. Asked at the end of each area:

- Is the area relevant?
- Is the mix of items relevant to the area? Some area that should be removed, or some item that is missing?

General about the test. Asked at the end of the interview:

- Is the mix of competency areas relevant for the purpose? Some area that should be removed, or some area that is missing?

Key check

For each question:

- Do you agree with the answer?

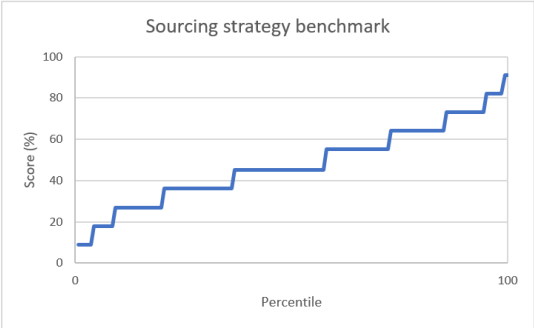
Distractor evaluation

For each question:

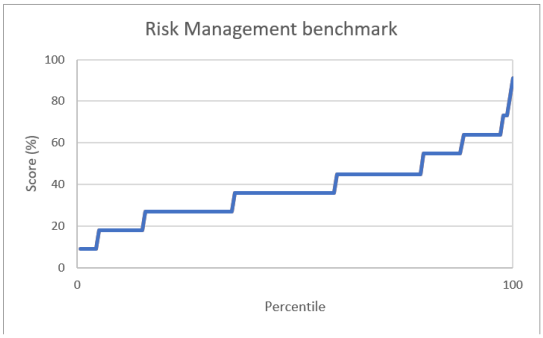
- Is there any other answer that could be correct?

Appendix B

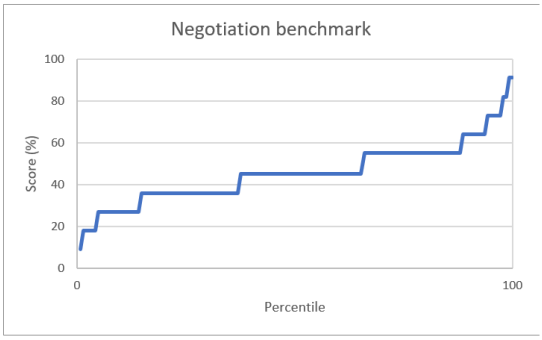
Empirical cumulative distribution for all competence areas



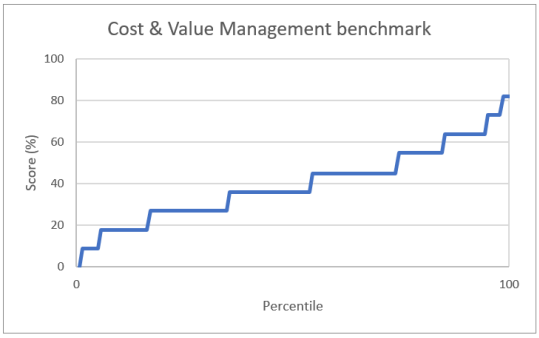
(a) Sourcing Strategy



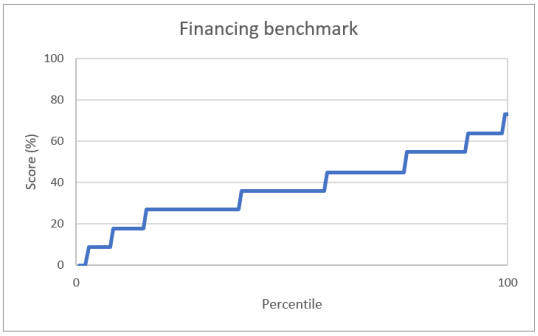
(b) Risk Management



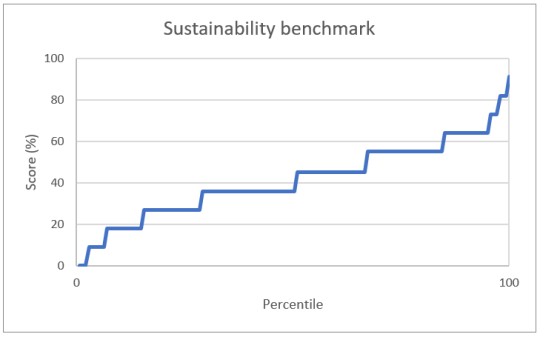
(c) Negotiation



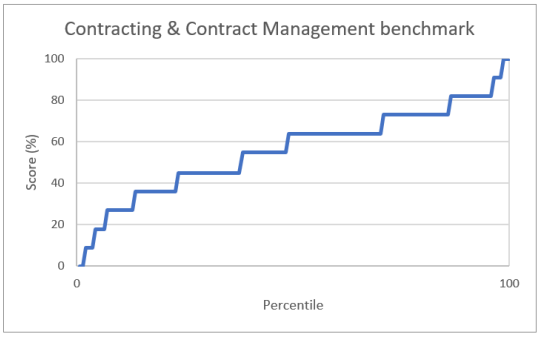
(d) Cost Value Management



(e) financing



(f) Sustainability

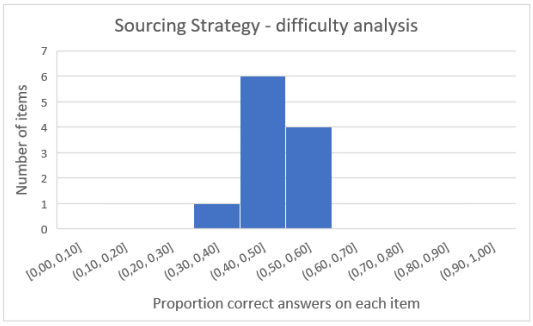


(g) financing

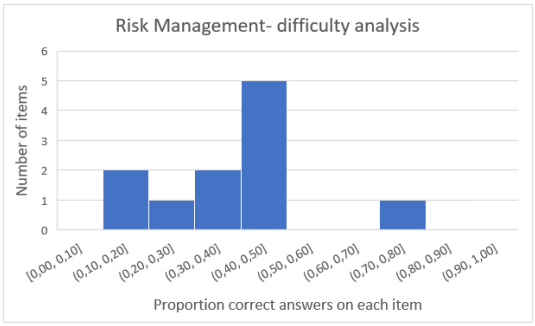
Figure B.1: Empirical cumulative distribution for each of the competency areas.

Appendix C

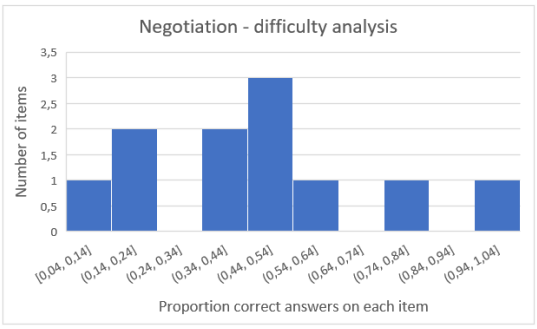
Difficulty distribution for all competence areas



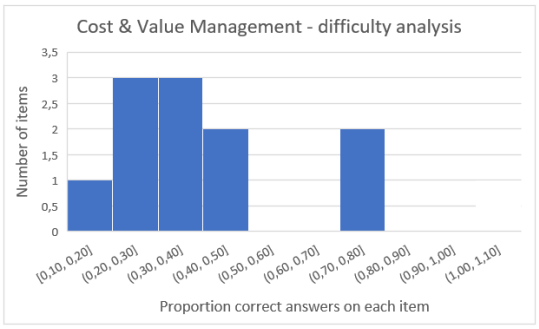
(a) Sourcing Strategy



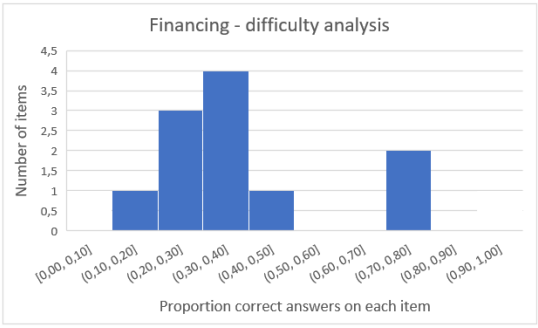
(b) Risk Management



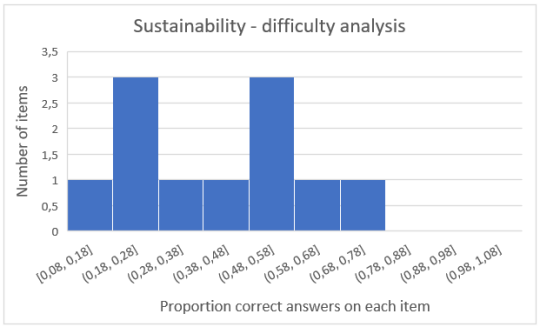
(c) Negotiation



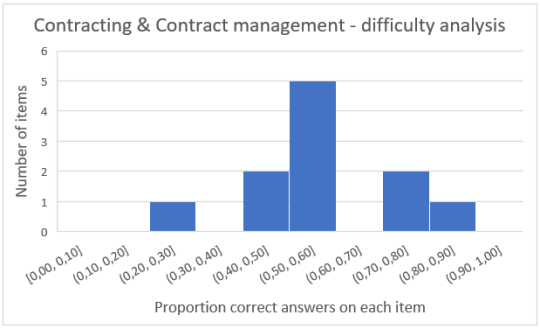
(d) Cost Value Management



(e) financing



(f) Sustainability



(g) financing

Figure C.1: Difficulty distribution for the seven areas.