# A SITUATIONAL JUDGEMENT TEST FOR ENGINEERS TO EVALUATE THEIR PROFESSIONAL STRENGTHS & WEAKNESSES

**Darren Carthy** [1]
Engineers Ireland
Dublin, Ireland
0000-0002-9810-1189

**Sofie Craps**
LESEC, KU Leuven
Leuven, Belgium

**Greet Langie**
LESEC, KU Leuven
Leuven, Belgium

**Kevin Gaughan**
TU Dublin
Dublin, Ireland

**Prof. Brian Bowe**
CREATE, TU Dublin
Dublin, Ireland

***Conference Key Areas***: *Engineering Skills, Assessment*
***Keywords***: *Skills, Competences, Graduate, Assessments*

## 1 INTRODUCTION

This paper reports on the development and evaluation of a 23 item Situational Judgement Test (SJT) with scenarios tailored to the engineering profession. The SJT was developed around the PREFER model, with the support of professional engineers and academics in 11 panel discussions. In total 53 engineering professionals and academics were consulted during the development of both the item stems and the item responses of the SJT. Subsequently, the SJT was rolled out to 334 final year and masters students enrolled in engineering programmes at TU Dublin and KU Leuven respectively. After taking part in the test, students were sent automated reports on their performance and the test which highlighted how their response compared to a response gathered from a professional engineer with feedback on how they might improve their competence in a particular area, while also commending their performance in other areas. The results of this study highlight that 8 SJT items had significantly lower mean scores when compared with the test-mean. These items, which were related to *perseverance, client focus, vision, planning and organising, solution orientation, team player, work organisation, clear communication* and *networking* all represent potential competence deficits in the population of final year and master students that were tested. This work adds to engineering education scholarship by providing an engineering-specific SJT that enables educators to identify areas of relative strength and weakness in students' professional judgements in order to better prepare them for their future careers.

### 1.1 Rationale

Over the past three decades, there has been a strong emphasis on improving the employability of engineering students in order to address the mismatch between graduate skills and labour market expectations in the field of engineering (1). A recent meta-analysis by Passow & Passow (1) discovered fifty-two articles published over the past three decades regarding the professional skills that engineering programmes should emphasise. Looking at just the past decade, a strong emphasis on professional skills in engineering programmes has resulted in the development of a multitude of learning resources, courses, interventions and assessments (2) that attempt to address students' lack of competence in these professional areas, and as alluded to in a previous paper, the vast majority of which rely on student self-assessment of their competence via pre and/or post intervention (2).

Despite three decades of research and intervention into students' professional skills, the skills mismatch for engineering professions in Ireland and Belgium is not reducing (3). Action needs to be taken to provide a more direct means of assessment of students professional competence, assessment that can also be used to provide prompt and directive feedback to students about how their professional competence can be improved. The type of assessment that brings us back to Boud's reframing of the purpose of assessment, which shifts the focus from viewing assessment as confirmation of the achievement of a particular learning outcome, to the use of assessment as a tool for informing judgement (4).

### 1.2 Literature review

The practice of including Situational Judgement Tests (SJT's) in the candidate selection process, particularly in job interviews, has become increasingly popular in recent years (5).This section presents the taxonomy of a SJT item, how SJTs are scored and how success in SJTs relate to job performance. SJT's have been used in psychological assessment for almost a century, with the first documented SJT appearing in the George Washington University Social Intelligence Test (6). The first section of the test was entitled J*udgement in Social Situations* which contained a number of social situations that presented problems, each followed by four possible solutions to that problem (6).

The items of an SJT are composed of two principal components (7).The first component is called the *item stem*, this is usually set in a professional work environment and involves a conversation between two or more actors. The contexts present a dilemma which is outlined in dialogue by one of the actors. The second component of an SJT item area the potential *item responses* that the second actor can provide to the first actor who presented the problem statement to attempt to address the issue.

There is substantial variation in how SJTs instruct the participant to select responses. Asking the participant for only a single most preferred response can result in faking, particularly in high stakes scenarios, and guessing or failure to engage thoughtfully with the test. A number of strategies have been developed to address this, one of which is to ask the participant to identify a best and worst response, forcing the respondent to reflect on why a response is appropriate or inappropriate rather than simply selecting the optimal response (7). This of course introduces ipsativity (8) to the test, in that a ranking is introduced to the responses. This can lead to issues with reliability analysis as the data collected are far less *granular* i.e a four-response SJT item scored Ipsatively has a theoretical maximum score of four. A more favourable method of *rating* responses or normative rating, can be employed instead and allows a more granular score to be attached to each item rating and so a four-response SJT item scored normatively on a 5 point Likert scale produces far more variation in score than an ipsatively scored item. This is a general issue faced in all Multiple Choice Question (MCQ) testing, but the use of normative scoring has particular significance to the scoring of SJT's as it allows more nuanced data to be collected from subject matter experts, who's responses may shape the scoring key of particular test items.

Patterson (9) carried out a systematic review of the use of SJTs in the evaluation of a number of non-cognitive factors including empathy, integrity and resilience. The review found that SJTs compared favourably with IQ tests and personality tests in predicting job success and represented a cost-effective means of candidate selection when compared with direct observation through structured interview. In the assessment of candidates' interpersonal skills Lievens (10) found "significant added value" in using SJTs over cognitive tests alone in predicting interpersonal skills. Motowidlo, Dunnette & Carter (11) found poor correlation between test scores and the GPA of participants but significant correlation with interpersonal skills r = .21 and negotiation r = .50 respectively, which were evaluated in interviews with test participants. An SJT developed by O'Connell et al.(12) shared variance with cognitive ability r = .33, conscientiousness r = .33 and agreeableness r = .31 which are established predictors of job success and the results are in good agreement with previous findings (13–15). The above literature seems to suggest that while SJTs are not predictors of academic success they are reliable predictors of future job performance when aligned with the appropriate professional skill.

### 1.3 Research objectives & question

In order to answer the research question: *"can a SJT be an effective means of assessing engineering students professional competence?"*, four research objectives were identified:

1) Develop an novel SJT with items tailored specifically to engineering professional scenarios.

2) Evaluate the SJT with all stakeholders including students, academics and industry professionals.

3) Identify the areas of professional strength and areas for improvement in the sample of engineers tested.

4) Provide students with prompt feedback on these strengths and areas for improvement.

## 2 METHODOLOGY

This research utilised a fixed mixed methods design, in which the quantitative and qualitative methods employed were predefined, and did not emerge from research findings in a previous stage of the research (16). The development of the SJT required the application of qualitative methods to evaluate respondents accounts about the content of the SJT items. The roll out of the SJT required the application of statistical tests to make inferences about the data collected from students who took part in this phase of research.

### 2.1 Methods

The method used in this study builds on a previous study of Craps et al (2021) who developed a validated Professional Roles Model for early career engineers (PREFER model) (17). A list of skills was developed by Binder Dijker Otte (BDO) (a consultancy with a division in Human capital) using a seminal piece of research by Bartram in which 29 validation studies (n= 4861) of his skills framework "the eight great professional skills" were meta-analysed (18). This list was brought to 13 expert panels in Belgium, Ireland and the Netherlands, all of whom employed engineers. Fifty-five panellists took part in the research; 47 male and eight female panellists who were predominantly engineers (44 engineers, 8 HR managers and 3 engineering managers with HR expertise) comprised the thirteen expert panels. In these panels, the competencies listed were mapped to the three professional roles outlined by the PREFER model, which describes three roles that early career engineers can take on when entering to the labour market: product leadership roles (focusing on radical innovation and research and development), operational excellence roles (focusing on product or process optimisation and increasing efficiency) and customer intimacy roles (focusing on tailored solutions forspecific clients).As well as these three role profiles, the outcome of these panels was a set of twenty-three skills and their descriptions. Once these had been identified, these skills were used as the basis for the development of the SJT items. The advantage of this approach was that items could be framed in a particular context by design, basing each item on one of the twenty-three skills that had been identified. Designing the items from scratch also allowed the situations to be kept to an appropriate length. More detailed questions result in higher validity but this must be tempered by keeping the cognitive loading of the items to a minimum (19).

In this study, once the items had been initially drafted, the test was reviewed by a further set of 11 panels, this time made up of academics and industry professionals alike. In total, 53 people took part in the panels; 33 males and 20 females. Three panels were academic, made up of lecturing staff from the schools of civil & structural engineering, school of mechanical engineering and school of electrical engineering at TU Dublin and KU Leuven respectively, who reviewed the item stems. The items stems were also reviewed by panels from industry, with ESBI, Siemens and ENGIE. Once the items stems had been reviewed and the feedback integrated, The item responses were scored by panels from ARUP, Siemens (2 panels), Bosch, and Materialize to generate a scoring key that would be used to compare with students responses to the items.

Once a revised draft of the SJT items had been created, the test was divided into 3, resulting in 3 tests with between 7-8 items each. It was decided to keep the items grouped by role, rather than randomly assigning items to each test. The tests were brought online using BDO's test platform and links were disseminated to 334 final year undergraduate engineering students and masters students at TU Dublin and KU Leuven The rationale for selecting final year and masters students was that they represented the students who were closest to joining the labour market. Through their potential exposure to work placements, internships, guest speakers and site visits it was posited that their responses should align well to the responses collected from the panels with industry, and where they did not align well, a mismatch could be identified in their competence. All students who took part in the rollout of the SJT received a feedback sheet that compared their responses on each item to the responses of industry professionals, to provide them with a means of reflecting on their competence. The feedback was sent automatically through the BDO online platform via pdf.

### 2.2 Qualitative data treatment

The first stage of the evaluation of the SJT began with a desktop review of the 23 test items by staff from TU Dublin and KU Leuven; their qualitative feedback was documented. In the second stage, the test was brought to three expert panels. These panels were comprised of junior engineers, senior engineers, engineering management and HR professionals from ESBI, Siemens and ENGIE. During these panels, the participants were asked to evaluate the item stems to check if the items were suitable representations of the professional skill which it had been related to while their qualitative remarks were recorded. In the third stage, the test was brought to three expert panels comprised of lecturing staff at TU Dublin with backgrounds in the engineering industry and psychology from the schools of Civil & Building Services engineering, Mechanical & Design Engineering and Electrical & Electronic Engineering. In these panels, the participants were asked to assign two or three skills to each SJT item to check for alignment between the item and the skill it was written to represent.The data from the third review stage were compiled and reviewed by the researcher. Following revision of the content of the test items a fourth review stage began where the four possible responses to the scenarios presented in each item were reviewed in a further four expert panels with practicing engineers, engineering management and HR professionals at ARUP, BOSCH, Materialise and two panels with Siemens. The participants were asked to indicate their level of experience and their role along with their scores of the level of appropriateness of *each* item response on a 1-5 Likert scale, their qualitative remarks were also recorded. The scores provided by the experts were compared to the theoretical scoring key established by the researchers to establish a hybrid scoring key.

### 2.3 Quantitative data treatment

At an item level, where scores on each of the four possible responses could be aggregated, providing a theoretical maximum score of 24, the data were found to be normally distributable, and parametric statistics were utilised (20). T-tests were carried out to look for differences in sample means between different items, in order to determine students relative strengths and weaknesses in their evaluation of the scenarios presented in the SJT items. Participantion in the research was voluntary and di not form part of the students final assessment for any module. This research was conducted with Ethical approval from the TU Dublin ethics committee (REC-17-112).

## 3 RESULTS

### 3.1 Box and Whisker plots

The distribution of the students' scores for each item are displayed in figures 1.1, 1.2 & 1.3 respectively, with the mean score represented by the black line on each box plot and the data which fall within the normal distribution represented by the shaded area. Data was tested for normality using the Shapiro-Wilks method and determined not to be significantly different from a normal distribution. Results of the t-tests are available in Appendix A.

As illustrated in Figure 1.1, *planning and organising, solution orientation, team player* and *work organisation* had lower mean scores than the other items. It was unsurprising that *planning and organising and work organisation* were of similarly low scores as the operationalisation of these items was very similar; in both instances a cognitively loaded item was avoided, for example an item where an optimised schedule had to be created – as this would not fit well within the taxonomy of an SJT item. Instead, responses outlining consistent and inconsistent plans were created and the respondent was asked to rate each of these in terms of their utility for completing a particular task.

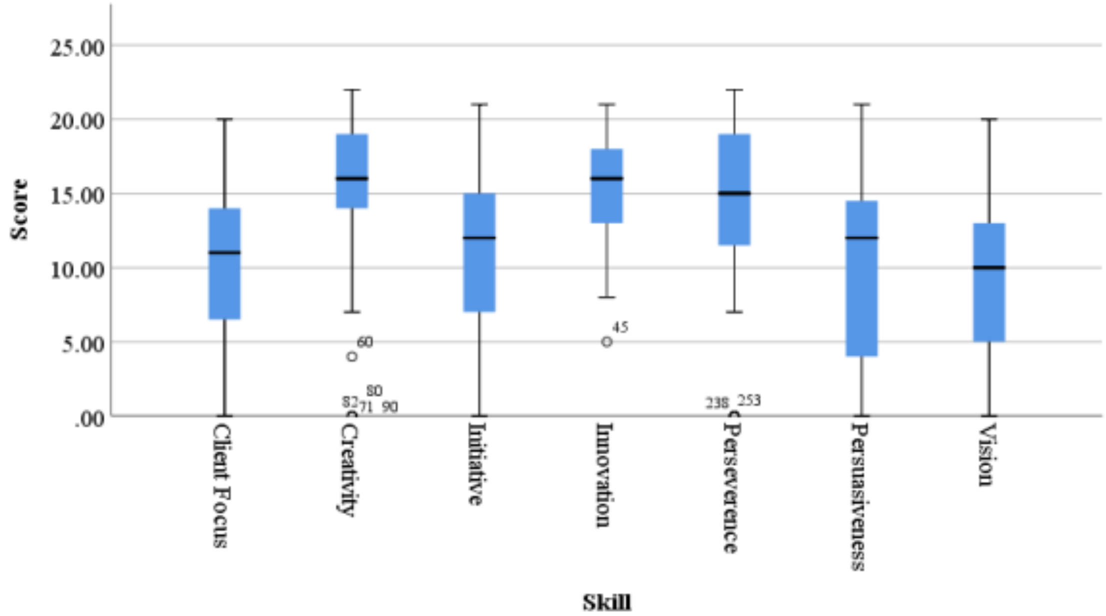


Figure 1.1 Box-Whisker plots for Product Leadership test

With reference to Figure 1.2, the *work organisation* item was of particular concern due to the low mean score, but also the variance in the score. The distribution of scores within the normal distribution presented as the shaded boxes in Figure 1.2 illustrate that the tail of the distribution of scores for *work organisation* was firmly placed between a score of 0-5. This was also the case for *planning and organising*, although to a slightly lesser extent These presented areas of weakness in the students' professional skills that should be addressed.

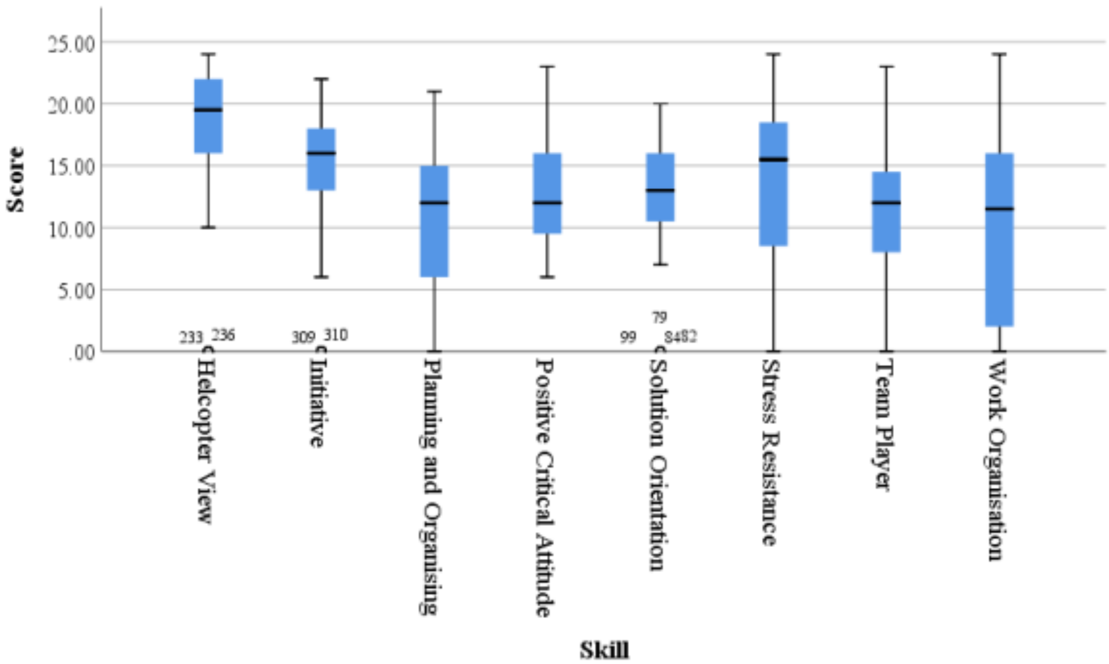


Figure 1.2 Box-Whisker plots for Operational Excellence test

With reference to Figure 1.3, *clear communication* and *networking* were found to have significantly lower mean scores than the test mean. The *networking* item was operationalised around a willingness to proactively network with an auditance of potential clients, or to stand back and take a more passive approach. This item may have been mediated by the introvert-extrovert personality trait, however it does suggest that on average, students are unwilling to engage proactively at a networking event.

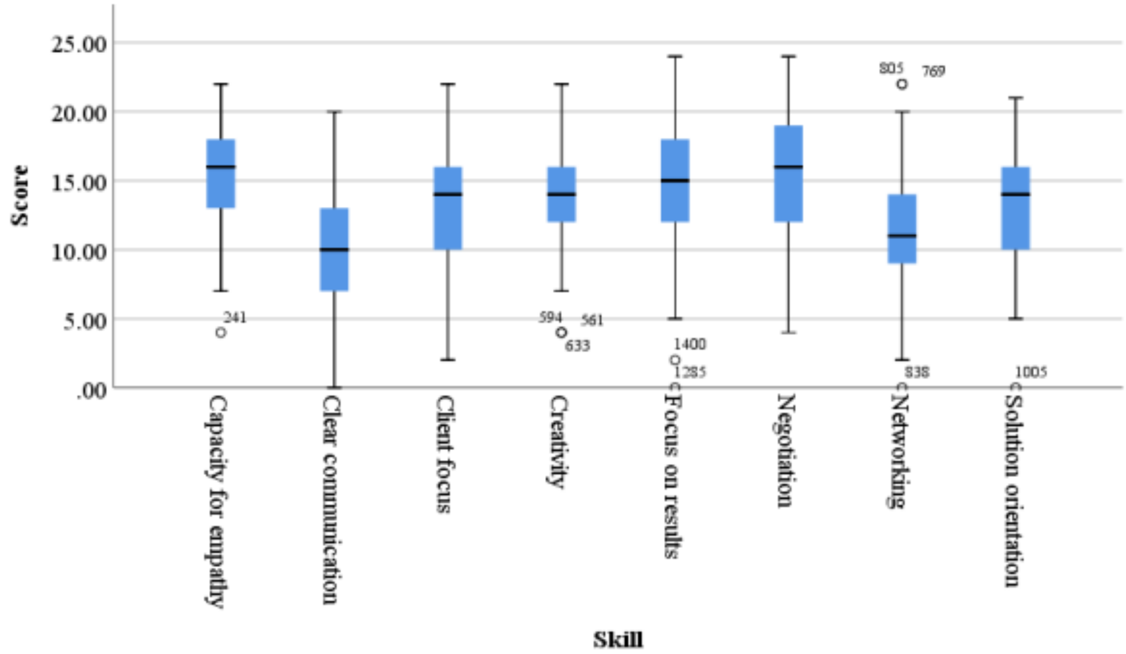


Figure 1.3 Box-Whisker plots for Customer Intimacy test

### 3.2 Paired sample t-tests

Items with the lowest mean score were identified and t-tested by comparing the item-mean to the test mean for each of the 3 tests, the results of these paired sample t-tests are presented in Appendix A. Eight of the twenty-three items of the SJT had significantly lower mean scores than the remaining thirteen items. These items, which were related to *perseverance*, *client focus*, *vision, planning and organising*, *solution orientation, team player, work organisation, clear communication* and *networking* all represent potential skill deficits in the population of final year and master students that were tested.

## 4 SUMMARY AND ACKNOWLEDGMENTS

As the students tested in this study are now in the labour market, it is useful to thinking about the ways in which these engineers may begin to bridge their skill gaps. As the sample of students tested are now in the labour market, implementation of the 70:20:10 model (21) of training may be useful, in which 70% of these gaps may be addressed through engagement with challenging projects, 20% engaging with a mentor to identify potential avenues, both formal and informal to bridge these gaps and 10% with further formal learning by engaging in courses offered by third level institutions and engineering professional bodies alike. Beyond the immediate needs of those students who were tested, further use of the SJT as a means of identifying skill gaps could be used proactively to inform engineering curriculua at third level, and also inform organisation's graduate development programmes regarding the types of training required of the incoming cohort of engineering talent.

This research was funded by the PREFER (Professional Roles and Future Employability of EngineeRs) project, a European commission funded research project under the Erasmus+ funding scheme (Reference: 2014-1-BE02- KA200-000462).

## REFERENCES

1. Passow HJ, Passow CH. What Competencies Should Undergraduate Engineering Programs Emphasize? A Systematic Review. J Eng Educ. 2017;106(3):475–526.
2. Carthy D, Gaughan K, Bowe B. Reconceptualising studies of engineering professional skills : Moving the field on. In: 47th SEFI Conference, 16-19 September 2019, Budapest, Hungary. 2019.
3. European Commission. Ireland : Mismatch priority occupations. 2016.
4. Boud D. Reframing assessment as if learning were important. In: Rethinking Assessment in Higher Education: Learning for the Longer Term. London Routledge; 2007. p. 14–26.
5. Chan D, Schmitt N. Situational Judgement Tests. In: The Blackwell Handbook of Personnel Selection. 2005. p. 219–42.
6. Hunt T. The measurement of social intelligence. J Appl Psychol. 1928;12(3):317–34.
7. Mcdaniel MA, Nguyen NT. Situational Judgment Tests : A Review of Practice and Constructs Assessed. Int J Sel Assess. 2001;9(June):103–13.
8. Hicks LE. Some properties of Ipsative, Normative and forced choice normative measures. Psychol Bull. 1970;74(3):167–84.
9. Patterson F, Ashworth V, Zibarras L, Coan P, Kerrin M, O'Neill P. Evaluations of situational judgement tests to assess non-academic attributes in selection. Med Educ. 2012;46(9):850–68.
10. Lievens F. Adjusting medical school admission: Assessing interpersonal skills using situational judgement tests. Med Educ. 2013;47(2):182–9.
11. Motowidlo SJ, Dunnette MD, Carter GW. An Alternative Selection Procedure : The Low-Fidelity Simulation. J Appl Psychol. 1990;75(6):640–7.
12. Connell MSO, Hartman NS, Mcdaniel MA, Lee W, Iii G, Lawrence A. Incremental Validity of Situational Judgment Tests for Task and Contextual

Job Performance. Int J Sel Assess. 2007;15(1).
13. Chan D, Schmitt N. Video-Based Versus Paper-and-Pencil Method of Assessment in Situational Judgment Tests : Subgroup Differences in Test Performance and Face Validity Perceptions. J Appl Psychol. 1997;(1):143–59.
14. Clevenger J, Pereira GM, Wiechmann D, Schmitt N, Harvey VS. Incremental Validity of Situational Judgment Tests. J Appl Psychol. 2001;86(3):410–7.
15. Whetzel DL, Mcdaniel MA, Nguyen NT. Subgroup Differences in Situational Judgment Test Performance : A Meta-Analysis. Hum Perform. 2008;21:291–309.
16. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research. SAGE Publications; 2018.
17. Craps S, Pinxten M, Knipprath H, Langie G. Different roles, different demands. A competency-based professional roles model for early career engineers, validated in industry and higher education. Eur J Eng Educ. 2022;47(1):144–63.
18. Bartram D. The great eight competencies: A criterion-centric approach to validation. J Appl Psychol. 2005;90(6):1185–203.
19. Mcdaniel MA, Morgeson FP, Finnegan EB, Campion MA, Braverman EP. Use of Situational Judgment Tests to Predict Job Performance : A Clarification of the Literature. J Appl Psychol. 2001;86(4):730–40.
20. Vidakovic B, Kvam PH. Chapter 6: Goodness of Fit. In: Nonparametric Statistics with Applications to Science and Engineering. Wiley; 2007.
21. Lombardo, Michael M; Eichinger RW. The Career Architect Development Planner 5th edition. Lominger International; 2000.

## APPENDIX A

| Item | Mean | Standard deviation | t | df | Sig |
|---|---|---|---|---|---|
| Product leadership | | | | | |
| Innovation | 3.356 | 4.788 | 5.384 | 58 | 0 |
| Creativity | 3.153 | 4.266 | 5.676 | 58 | 0 |
| Vision | -3.034 | 4.017 | -5.801 | 58 | 0 |
| Persuasiveness | -1.932 | 3.81 | -3.896 | 58 | 0 |
| Perseverance | 1.661 | 3.646 | 3.499 | 58 | 0.001 |
| Initiative | -1.034 | 3.429 | -2.316 | 58 | 0.024 |
| Client focus | -1.814 | 3.457 | -4.03 | 58 | 0 |
| Operational excellence | | | | | |
| Positive critical attitude | -0.1039 | 5.988 | 1.255 | 76 | 0.879 |
| Solution orientation | -0.0649 | 4.072 | 0.859 | 76 | 0.889 |
| Team player | -1.662 | 4.287 | -0.689 | 76 | 0.001 |
| Helicopter view | 4.402 | 4.139 | 5.342 | 76 | 0 |
| Initiative | 1.324 | 3.529 | 2.125 | 76 | 0.002 |
| Work organisation | -2.389 | 5.373 | -1.17 | 76 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| Stress resistance | 0.441 | 5.053 | 1.588 | 76 | 0.446 |
| Planning and organising | -2.363 | 4.032 | -1.448 | 76 | 0 |
| Customer Intimacy | | | | | |
| Client focus | 0.848 | 5.099 | 1.55 | 204 | 0.018 |
| Capacity for empathy | 1.921 | 3.907 | 2.46 | 204 | 0 |
| Clear communication | -3.448 | 4.179 | -2.873 | 204 | 0 |
| Creativity | 0.307 | 3.141 | 0.739 | 204 | 0.163 |
| Networking | -2.224 | 3.453 | -1.748 | 204 | 0 |
| Solution orientation | -0.326 | 3.46 | 0.149 | 204 | 0.178 |
| Negotiation | 1.531 | 4.304 | 2.124 | 204 | 0 |
| Focus on results | 0.97 | 3.725 | 1.483 | 204 | 0 |

Table 1.1 Paired sample t-tests for each item, separated by test