

# Validation of the Inquiry Skills Test

*Are Inquiry Skills related to Science Motivation and Understanding of a Science Text?*

Katja Hoffmann, s1007831

University of Twente

Faculty for Behavioral Science, Psychology

Department of Instructional Technology

Examination Committee

Dr. P. Wilhelm

Dr. T.H.S. Eysink

University of Twente

Enschede, 14 december 2012

## **Abstract**

Goal of this study was to provide evidence for the validity of the Inquiry Skills Test (IST; Horstink, 2005). The IST and the Dutch Science Motivation Questionnaire (DSMQ; Wilhelm, Stellmacher, Eysink, ten Klooster, submitted) were given to a sample of 44 eleventh-graders (pre-university education) and understanding of a science text was tested. Reliability of the instruments was investigated and compared to earlier studies on the IST. Correlation analysis was performed.

The IST had, except for two subtests, a high reliability. The IST scores correlated with the understanding of the science text. This supports that the IST does measure science skills. No correlation was found between IST and DSMQ scores. However, science motivation and text understanding correlated. A regression analysis showed that the IST scores, contrary to the DSMQ scores, significantly predicted the scores on the text understanding assignment. This provides additional evidence for the validity of the IST.

Summarized the findings support the validity of the interpretation of the IST scores. Further studies should be performed with a sample that has no science experience to make the results more generalizable. Also a factor analysis is recommended to test whether all items measure the skills they are expected to measure.

## **Abstract**

Doel van deze studie was de validiteit van de Inquiry Skills Test (IST; Horstink, 2005) te ondersteunen. Bij een steekproef van 44 5 VWO leerlingen werden de IST en de Dutch Science Motivation Questionnaire (DSMQ; Wilhelm, Stellmacher, Eysink, ten Klooster, ingediend) afgenomen. Verder werd het begrip van een wetenschappelijke tekst getest. De betrouwbaarheid van de instrumenten werd getoetst en met de resultaten uit vroegere studies met de IST vergeleken. Correlatie analyses werden uitgevoerd.

Met uitzondering van twee subtesten, had de IST een hoge betrouwbaarheid. Er werd een correlatie tussen de IST scores en het begrip van wetenschappelijke teksten gevonden. Dit ondersteunt de hypothese dat de IST onderzoek vaardigheden meet. Geen correlatie werd gevonden tussen de IST en de DSMQ scores. Echter werd tussen onderzoek motivatie en begrip van wetenschappelijke teksten een correlatie gevonden. De regressie analyse liet zien dat de IST scores, in tegenstelling tot de DSMQ scores, voorspelende waarde op de scores op de wetenschappelijke tekst hadden. Dit is een aanvullend ondersteuning voor de validiteit van de IST.

Sammengevat ondersteunden de resultaten de validiteit van de interpretatie van de IST scores. Vervolgonderzoek zou met een steekproef moeten worden uitgevoerd die geen ervaring met wetenschap heeft om de resultaten beter te kunnen generaliseren. Verder wordt een factoranalyse aangeraden om te testen of alle vragen de vaardigheden meten die zij geacht worden te meten.

## **Index**

|  |           |
|--|-----------|
| <b>Introduction</b> .....  | <b>1</b>  |
| <i>The Inquiry Skills Test</i> .....   | 3         |
| <i>Other validation research</i> .....   | 7         |
| <i>Validation of the IST</i> .....   | 9         |
| <i>Inquiry skills and science motivation</i> .....   | 10        |
| <i>Inquiry skills and understanding of a scientific text</i> .....                             | 10        |
| <i>Summary: research question and hypothesis</i> .....   | 11        |
| <b>Method</b> .....  | <b>12</b> |
| <i>Respondents</i> .....   | 12        |
| <i>Instruments</i> .....   | 12        |
| <i>Inquiry Skills Test</i> .....   | 12        |
| <i>Dutch Science Motivation Questionnaire</i> .....  | 14        |
| <i>Scientific text understanding assignment</i> .....  | 15        |
| <i>Procedure</i> .....   | 16        |
| <i>Scoring</i> .....   | 18        |
| <i>Data analysis</i> .....   | 19        |
| <b>Results</b> .....   | <b>20</b> |
| <i>Descriptive Statistics</i> .....  | 20        |
| <i>Group differences</i> .....   | 20        |
| <i>Internal Consistency</i> .....  | 21        |
| <i>Correlational analysis on total scores</i> .....  | 21        |
| <i>Subscale correlation between IST and the scientific text understanding assignment</i> ..... | 22        |
| <i>Correlation between IST total score and DSMQ sub constructs</i> .....                       | 24        |
| <i>Multiple linear regression analysis</i> .....   | 24        |
| <b>Conclusion and Discussion</b> .....   | <b>25</b> |
| <b>References</b> .....  | <b>30</b> |
| <b>Appendix</b> .....  | <b>36</b> |

## **Introduction**

Over the past years, the topic of inquiry learning has become more and more important in education. In 2003, Judith Lechner and Boris Wanders (De geschiedenis van het Technasium, n.d.) developed the idea for a new educational branch in the Netherlands, the Technasium. The idea was to offer learners in pre-university education the possibility to develop their research and design skills. The Ministry of Education supported the idea and in 2009, 41 schools had already joined the project. Many of the schools were awarded for their remarkable contribution to the development of science interest and science skills in Dutch learners.

Although the ambition to support science skills in Dutch education is relatively new, the general idea of inquiry learning has a long history. In his book “Logic – Theory of Inquiry” Dewey (1938) presents his idea of inquiry learning. He states that learning is always related to the context. Learners have to actively engage into situations where they can gain knowledge and have to interact with the world to gain knowledge from it (Dewey, 1910). This process follows five steps (Dewey, 1933). First, an occasion evokes an internal confusion the learner wants to resolve. In the second step, the confusion is specified based on prior knowledge and self-selected information. The goal is to determine which data or information is relevant for the solution of the problem involved. This leads to the third step in which hypotheses about possible solutions for the problem are constructed, a process that demands creativity. In the fourth step, new data is gathered based on experimentation. This is an active process with the goal to either discard or support the chosen hypotheses. In the final step, it is clarified whether the expectations are met. This includes conclusions and interpretations, based on the results.

Wells (1999) has a similar concept of learning. Based on Vygotsky’s theory about knowledge acquisition (Vygotsky, 1981; Vygotsky, 1987), Wells outlines the growing importance of constructivism in education. Vygotsky, who inspired the “social constructivist” theory, aimed in particular at more effective knowledge acquisition in social interaction. Following Vygotsky’s argumentation, Wells states that a classroom should work as a “Community of Inquiry” in which the learners use their environment to engage in an active search for knowledge.

Today, the topic of inquiry learning is still present. Studies have been based on Dewey’s theory of inquiry learning (e.g. Scalon, Anastopoulou, Kerawalla & Mulholland, 2011). Scalon et al. tested in how far their software “nQuire” supports the inquiry process. They found that nQuire supports learners as well as teachers during all steps of the inquiry

cycle. Additionally, they mention the concept of personal inquiry, where learners engage in inquiry processes on topics that interest them. The researchers highlight the possibility to extend inquiry processes outside the school curriculum. The study not only shows that inquiry skills are still relevant today, but also demonstrates how inquiry learning can be improved by today's technology. Moreover, Scalon et al. demonstrate that it is important to find out if learners and teachers need support for parts of the inquiry process.

Horstink (2005) sees a growing relevance of inquiry learning in the Dutch educational system and the need of Dutch teachers and learners to be supported in the inquiry process. She states that there has been a change in education in the Netherlands especially over the past 10 years from a teacher centered approach to a learner centered approach which emphasizes scientific reasoning and inquiry, like Dewey (1910; 1933; 1938) and Wells (1999) describe it. She cites a report of the Dutch Organization for Scientific Research (NWO [De Nederlandse Organisatie voor Wetenschappelijk Onderzoek], 2003). The NWO is a Dutch funding organization, which has the goal to promote science innovations and projects and to support knowledge exchange about educational research in the Netherlands. Every four years the NWO publishes a report, with the goal to plan the educational research topics in the Netherlands for the upcoming years to improve Dutch education. The report promotes ways of learning in which learners use inquiry to answer overall questions. Contrary to Dewey, however, the NWO highlights the importance of the teacher who dictates specific topics. This is important to make sure all learners in the Netherlands follow similar curricula. Based on this, Horstink recognized the value of an instrument that would enable teachers to detect differences in learners' inquiry skills. Using this knowledge, teachers could adapt their teaching strategies to promote inquiry skills in a way that fits every learner. For this purpose, Horstink developed the Dutch Inquiry Skills Test (IST; Horstink, 2005). Her goal was to develop an instrument that measures inquiry skills in tenth-graders (pre-university education) and up. This target group was chosen, because their curriculum prepares for eventual science careers for which they need inquiry skills.

Seven years later, the IST still has great relevance. The current educational research program of the NWO (2012) reports three important topics for the period 2012-2015: the development of domain specific higher order skills (e.g. science skills, design, problem solving, etc.), affective and motivational aspects of learning, and adaptive education. Obviously, science skills still play an important role in Dutch educational research. Special attention is paid to the necessity to detect differences in abilities between learners, because different age and competency groups need different teaching strategies (NWO, 2012). In the

report, the need for measurement methods for different skills is stressed. Goal of this study is therefore to validate the IST. As a valid measure for inquiry skills, it could be used to improve pre-university education.

### *The Inquiry Skills Test*

Horstink (2005) based the development of the IST on a detailed analysis of research on inquiry learning and inquiry skills. The theories of Dewey (1910; 1933; 1938) and Wells (1999) as described above formed the foundation for her definition. Therefore, the instrument should relate to the inquiry learning steps as described by Dewey (1933), namely *internal confusion, specification of the confusion, construction of hypothesis, experimentation and analysis of the results*. With regard to Wells, Horstink highlighted that the instrument should show whether learners understand how to interact with their environment to gain knowledge. Moreover, she designed the instrument for school purpose, where teacher interaction is important. Based on this foundation, Horstink decided to elaborate the definition of inquiry skills and search for existing and related tests she could use to construct a valid test for inquiry skills for her purpose and target group.

Horstink started with Bonstetters' (1998) theory on levels of inquiry learning. According to this theory, the levels of inquiry differ by how autonomous a learner performs a task. He describes five levels: *Traditional Hands-on Science Experiences, Structured Science Experiences, Guided Inquiry, Student Directed Inquiry* and *Student Research*. These levels range from much teacher guidance to much learner guidance. Horstink decided to base her instrument on Bonstetter's *Student Directed Inquiry* level, at which the teacher only influences the topic. This is realistic in school settings, because some guidance with regard to learning topics is necessary to stay in line with the curriculum. Similarly, Chinn and Malhotra (2002) differentiate between authentic science inquiry, which is directed by the learner and simple inquiry tasks, which are guided by the teacher.

Hereafter, Horstink elaborated on the definition of inquiry learning based on similar concepts like critical thinking, scientific reasoning and scientific discovery learning. By comparing these concepts with Dewey's and Wells' definitions of inquiry learning, Horstink narrowed down the skills that should be tested with the instrument. For example, Schaferman's (1991) definition of critical thinking closely resembles Dewey's inquiry learning definition. According to Kuhn (1993), scientific reasoning skills and inquiry skills have much in accordance and Reid, Zhang and Chinn (2003) see scientific discovery learning as a form of constructivist learning in which the learner engages in problem solving activities

with a scientific working style; this resembles Wells' theory of inquiry learning. Knowing that the definition of these skills resembles the definition of inquiry skills, Horstink could use tests that measure these skills to construct her instrument.

Next, Horstink had to decide how to test inquiry skills. Stokking and van der Schaaf (1999) differentiate between cognitive activities and research steps, which are interrelated but not equal. A cognitive ability, in this study an inquiry skill, can be necessary for different research steps and one research step, for instance designing an experiment, can call upon different cognitive abilities. Stokking and van der Schaaf state that in work and school settings it is simpler to focus on research steps. For her instrument, Horstink therefore adopted the view that science skills correspond to steps that have to be taken in order to do research.

The goal was to develop a domain independent test for inquiry learning. However, Stokking and van der Schaaf stress that inquiry skills are gained in situations where domain dependent knowledge plays a role (i.e., in practical experiments). Skills are domain dependent and skills needed in two different domains are never exactly the same. Therefore, Horstink takes into account that it is not possible to create a completely abstract test for inquiry skills.

De Jong and Njoo (in de Jong & van Joolingen, 1998) make a distinction between transformative processes (hypothesis, design experiments, etc.) which contribute to knowledge gain and regulative processes (e.g., planning or monitoring) that direct the research process. According to Horstink, the transformative processes are more in line with inquiry steps, because regulative processes can also play a role in other forms of learning.

The most important inquiry steps Horstink determined based on the analysis of literature were a) *definition of variables*, b) *construction of hypothesis*, c) *designing of experiments*, and d) *evaluation of the results and drawing conclusions*. The IST should measure skills pertaining to these steps in order to provide a complete picture of the level of inquiry skills.

According to Lavinghouzes (1997), inquiry skills can best be measured in laboratory experiments. Horstink however, wanted to develop an instrument to measure inquiry skills that is less time consuming and can be used within the school context. The IST should be a domain independent test and a not-curriculum related test. Horstink chose seven instruments and judged their usefulness for her own instrument. Based on literature search, Horstink selected instruments that measure abilities (e.g., critical thinking, science process skills) that resembled her definition of inquiry learning. Horstink judged the instruments based on five criteria: 1) connection to the four chosen inquiry skills, 2) suitability for the target group, 3) measurement goal (determine the level of inquiry learning), 4) duration, and 5) reliability



(Cronbach's Alpha of at least 0.70) and validity. The first criterion was the most important for Horstink. She analyzed the following tests:

**Watson-Glaser Kritisch Denken Test (WGKDT, Van Zanten, Dekker & Berkhout,, 1997):** The WGKDT is based on the WGCTA (Watson-Glaser Critical Thinking Appraisal, Watson & Glaser, 1964). The WGCTA contains five subtests: *Inferences, Recognition of Assumptions, Deduction, Interpretation* and *Evaluation of Arguments* (Watson & Glaser, 1994).

**California Critical Thinking Disposition Inventory (CCTDI, Facione & Facione, 1992) & California Critical Thinking Skills Test (CCTST, Facione & Facione, 1990):** Both tests were developed by Facione and Facione in the 90's. Whereas the CCTCI concentrates more on a general disposition to critical thinking, the CCTST tests skills and abilities with regard to inquiry learning. The CCTCI subscales *Open-mindedness, Analyticity, Cognitive Maturity, Truth Seeking, Systematicity, Inquisitiveness* and *Self-confidence*. The subscales of the CCTST are *Interpretation, Analysis, Evaluation, Inference* and *Explanation*.

**Test of Enquiry Skills (TES, Faser, 1979):** The TES is a domain dependent test (e.g. for the fields of natural science or history). The goal of the TES is to measure individual learning and inquiry learning.

**The Integrated Process Skills Test II (TIPSII, Okey, Wise & Burns, 1982):** The first version of the TIPSII, the TIPS, was developed as a non-curriculum related test (Okey & Dillashaw, 1980). Both versions contain 36 multiple-choice items. The second version was developed as a completely new test with similar items on the same difficulty level (Burns, Okey & Wise, 1985). The goal was to provide an alternative choice between two equal tests for process skills. Both tests have five subscales: *Identify Variables, Stating Hypothesis, Operational Definitions, Designing Investigations* and *Graphing and Interpreting data*.

**Cornell Critical Thinking Test (CCTT, Ennis & Millman, 1985):** The CCTT tests two levels, *Level X* and *Level Z*. X contains items on induction and deduction whereas Z consists of items about definitions and experiment plans.

**Critical Reasoning Test (CRT, Smith & Whetton, 1992):** This test contains a verbal and a non-verbal part and is divided in three subtests: *Analysis of Information*, *Evaluation* and *Assumptions*.

Based on the analysis, it was decided to construct the inquiry skills test with the TIPSII and the WGKDT. With regard to the content, the CCTST seemed like a better candidate than the TIPSII, but due to its long testing duration Horstink decided not to use it. The TIPSII does not test the fourth chosen skill *evaluation of the results and drawing conclusions*. Therefore the subtests *Conclusion* (i.e., *Deduction*) and *Interpretation* from the WGKDT were included, too. The Cronbach's alpha of the TIPSII is .86 and its completion takes 30 to 35 minutes (Okey, Wise & Burns, 1982). The Cronbach's alpha of the whole WGCTA is .81 and it can be finished within 40 minutes (Van Zanten, Dekker & Berkhout, 1997). Horstink translated the TIPSII into Dutch and used the already existing Dutch version of the WGCTA. She used all 36 items of the TIPSII, including twelve items for the subscale *Identify Variables*, nine for the subscale *Stating Hypothesis*, six for *Operational Definitions*, three for *Designing Investigations* and six for *Graphing and Interpreting Data*. Although the TIPSII subtest *Graphic and Interpreting Data* seemingly tests Horstink's inquiry learning step *evaluation of the results and drawing conclusions* this is not the case. The subtest *Graphing and Interpreting Data* subtests does only measure in how far learners can understand and organize outcomes of experiments. The WGKDT subtests *Conclusion* and *Interpretation* test in how far learners can determine the relevance of outcomes to answer a research question. Both subtests, *Conclusion* and *Interpretation*, consist of 16 items. In total, the IST contains 36 items from the TIPSII and 32 from the WGCTA.

Horstink used a cognitive capacities test (CCT) and an inquiry task to validate the IST. Based on research that showed a relation between cognitive capacities like numerical series and performance in an inquiry skills task (Veenman, 2004, Wilhelm, 2001), Horstink hypothesized that IST scores correlate with CCT scores. The CCT contains subtests for *Vocabulary*, *Linear Syllogisms*, *Word Analogies*, *Numerical Series* and *Hidden Figures*. For the inquiry task the FILE software was used (Flexible Inquiry Learning Environment, Hulshof, Wilhelm, Beishuizen & van Rijn, 2005). With this program, abstract and domain dependent inquiry tasks can be constructed and performed. Participants have to select variables to perform experiments and discover the model describing the relationships between the independent and the dependent variables in a task. Horstink used an abstract inquiry task. A positive correlation between IST scores and inquiry task scores was expected.

Twenty-four students participated in the study. The IST had a Cronbach's alpha of .68, with an alpha of .68 for the TIPSII related part and an alpha of .30 for the WGKDT subtests. Except for the TIPSII subtest *Identify Variables* ( $\alpha = .75$ ), the subtests had relatively low alphas (the alphas ranged from -.47 to .75). Using the summed score, it appeared that IST scores were related to inquiry learning performance ( $r = .62, p < .01$ ), whereas no correlation with the CCT was found ( $r = .24, p = .26$ ). The correlation between the TIPSII score and the inquiry task performance was increased with 0.042 through the WGKDT subtest scores. There was also a significant correlation between inquiry task performance and the CCT scores ( $r = .50, p < .01$ ). A regression analysis revealed that the IST and the CCT explained 44.9% of the variance in the performance on the inquiry task scores. However, only the IST was a significant predictor ( $\beta = .56, p < .01$ ). The IST explained 37.9% of the variance in the inquiry task scores.

The hypothesis that the IST scores and the inquiry tasks scores should correlate was confirmed. This supported the notion that the IST does measure inquiry skills. However, the hypothesis that inquiry skills and cognitive capacities correlate was not supported. Horstink recommended repeating the study with bigger samples and learners of different ages and from different schools, to make it possible to generalize the results to other samples and to other institutions and organizations. Moreover, she suggests using a concrete learning task, to test whether the IST could also predict inquiry learning performance in a content-rich domain.

#### *Other validation research*

After the initial validation, several other studies were performed to test whether the IST is a reliable and valid instrument to measure inquiry skills. The first study was done in 2007 (von Ruedorffer, Streese, Kamps & Schmitt, 2007). Goal of this study was to heighten the internal consistency of the instrument by enhancing the number of TIPSII items. The researchers added 16 items, with the goal that each subscale would contain at least 10 items. Additionally, the openness questionnaire from the NEO-PI-R was included. It was hypothesized that inquiry skills would correlate with openness, because in an earlier study competences that play a role in inquiry learning correlated with autonomy (Schönrock-Adema, 2002). According to von Ruedorffer et al., autonomy relates to the openness concept of the Big Five. The study also included the CCT used before and an abstract inquiry task from FILE, to replicate Horstink's study with the extended questionnaire and a different sample. A sample of 30 university students with a mean age of 20.1 ( $SD = 1.4$ ) was used. The reliability of the total scale rose to  $\alpha = .84$ . Overall, the Cronbach's alpha on all subscales increased. No

significant correlation was found between the IST scores and the cognitive capacities ( $r = .39$ ,  $p = .06$ ). The same applied to the IST scores and the openness scores ( $r = .01$ ,  $p = .96$ ). The IST and the inquiry task correlated high ( $r = .66$ ,  $p < .01$ ). A multiple linear regression analysis revealed that the IST and the CCT explained 44% of the variance in learning performance. However, only the IST contributed significantly to the explanation of variance ( $\beta = .19$ ,  $p < .01$ ). The IST predicted 43.9% of the variance in scores of the inquiry learning task.

Geerdink, Rijken and Vennemann (2009) also conducted a validation study. The IST, the CCT and an inquiry learning task were used. The study included 23 students, 13 students were from another study on the IST (Hensel, Kuipers and Laseur, 2009). The mean age was 22.2 years ( $SD = 2.00$ ). A high inter-item correlation was found for the IST, the TIPSII related part and the WGKDT related part. The coefficients were respectively  $\alpha = .84$ ,  $\alpha = .85$  and  $\alpha = .60$ . Again the Cronbach's alpha for the WGKDT related part was only moderate. An analysis of the results revealed a low but significant correlation between the IST and performance on the learning task ( $r = .51$ ,  $p = .02$ ) and a strong significant correlation between the IST and the CCT ( $r = .69$ ,  $p < .01$ ). Despite the results of the correlation analysis the IST and the CCT both had no significant predictive effect on the domain dependent learning task.

Another study was carried out by Kip, Looge, Fens and Heilema (2009). They also used the IST, the CCT and a concrete learning task from FILE. Additionally, they included the Amsterdam Job Interest Questionnaire (Amsterdamse Beroeps Interesse Vragenlijst, ABIV; Evers, 1992). They hypothesized a relation would exist between learners' interest in a scientific job and the level of inquiry skills. Twenty tenth-graders from pre-university education filled in the IST and the ABIV with a mean age of 15.5 ( $SD = 0.67$ ). The Cronbach's alpha of the IST was .72. For the TIPSII related part again a high coefficient was found ( $\alpha = .84$ ), whereas the WGKDT related parts *Conclusion* and *Interpretation* had a low internal consistency,  $\alpha = .27$  and  $\alpha = .20$  respectively. No correlation was found between inquiry skills and job interest on the three subscales with which a correlation was expected: *Exact-Scientific* ( $r = .31$ ,  $p = .19$ ), *Alpha-Scientific* ( $r = .08$ ,  $p = .74$ ), and *Social-Scientific* ( $r = .28$ ,  $p = .23$ ). So, no support was found that learners with good research skills are also interested in doing research in their job.

A similar study was conducted one year later (Lange, von der Goltz, & Drawert, 2010). In this study, it was tested whether IST scores are related to CCT scores and/or scores on the Dutch Science Motivation Questionnaire (DSMQ; Wilhelm, Stellmacher, Eysink, ten

Klooster, submitted). Moreover, it was examined if there are gender differences on the IST. The study was carried out with 18 students. The students were on average 21.1 years ( $SD = 2.00$ ). The DSMQ and IST had high Cronbach's alphas ( $\alpha = .84/.89$ ). The TIPSII related subscales had moderate to high alphas, whereas alphas of the WGKDT related subtests *Conclusion* and *Interpretation* were again lower than .05 (i.e., .37, resp. .45). No significant correlation between the IST and the DSMQ was found ( $r = -.07$ ). A significant correlation was found between the IST and the CCT ( $r = .70, p < .01$ ). Moreover, a significant difference between men and women on the IST was found. Men scored higher on de IST ( $t = .02, p < .01$ ). The researchers suggested that small sample size and personal experiences might have explained their findings. All participants were university students; therefore possible negative experiences with research in the first study year could explain the low correlation between DSMQ scores and IST scores. A study with another, bigger sample was recommended.

#### *Validation of the IST*

Goal of this study was to validate the IST. In earlier studies on the validity of the IST it was tested whether inquiry skills correlated with cognitive capacities, scores on abstract and concrete inquiry tasks, age, job interests, and science motivation. The relation between IST scores, cognitive capacities, and scores on inquiry tasks was often studied, with varying results. Whereas Horstink (2005) and von Rueddorffer et al. (2007) found no correlation between IST scores and CCT scores, Geerdink et al. (2009) and Lange et al. (2010) did find such a relation. The same was true for the inquiry task. Horstink (2005) and von Ruedorffer et al. (2007) found a correlation and Geerdink et al. (2009) found no correlation with the IST scores.

In these studies it is hypothesized that the IST measures skills, which also play a role in the inquiry task. Abstract as well as concrete inquiry tasks have been part of most validation studies so far, therefore the inquiry task should be replaced by a new task in this study. A correlation between the IST scores and the scores of the new task would be an additional support that the IST measures the skills it is expected to measure. Additionally, no correlation has so far been found between the IST scores and science motivation or interest. However, the studies that hypothesized such a correlation had, due to various limitations, no ecological validity. Therefore, this hypothesis should be tested again. Thereby, the goal is to give additional support for the interpretation of the IST scores. As a valid instrument, the IST could be used in schools as intended by Horstink.

*Inquiry skills and science motivation.* In this study, two additional instruments are included to validate the interpretation of IST scores. The first is the Dutch Science Motivation Questionnaire (DSMQ; Wilhelm, Stellmacher, Eysink, ten Klooster, submitted). Science motivation and science skills are considered to be interrelated (Facione, 2000; Zusho, Pintrich & Coppola, 2003; Glynn & Koballa, 2006; Bianco, Higgins & Klum, 2009). The hypothesis that critical thinking and science disposition belong together was formulated by Facione in 2000. He claimed that with regard to critical thinking not only abilities are important, but that personal dispositions also count. An instrument to measure inquiry skills should therefore be related to a construct that measures the disposition to do science, like science motivation as measured with the DSMQ.

Zusho et al. (2003) showed that motivation plays a role in domain dependent science projects. It was found that the motivational aspects of self-efficacy and task value correlated with achievement in a chemistry course. This study supports that science motivation correlates with science skills. The DSMQ scores should therefore correlate with IST scores.

Glynn and Koballa (2006) found that motivation directs students' behavior. Students who were more motivated attained good results on science projects. They cite Brophy (1988) who defines motivation as "a student tendency to find academic activities meaningful and worthwhile and to try to drive the intended academic benefits from them" (pp. 205-206). Moreover, they differentiate between aspects of motivation like intrinsic and extrinsic motivation, goal orientation, self-determination, self-efficacy (Bandura, 1997), and anxiety. Glynn and Koballa (2007) developed the Science Motivation Questionnaire on which Wilhelm et al. (2010) based their construction of the Dutch Science Motivation Questionnaire. Glynn and Koballa showed that science success and science motivation correlate. It is expected that the IST measures inquiry skills of learners like a grade for a science project. Therefore, it is reasonable to assume that the IST scores also correlate with science motivation score. Such a finding would support the validity of the IST.

*Inquiry skills and understanding of a scientific text.* For this study, a scientific text understanding assignment was developed. The understanding assignment contained a scientific paper and questions about the paper. Learners had to answer questions about the text, find mistakes in the text and in the research design, and suggest possible corrections. All questions and assignments were based on the IST and its subtests. The skills that the IST measures are therefore thought to play a role in the understanding of the scientific text.

Both the IST and the text understanding assignment are domain independent inquiry tasks. As mentioned above the scientific text understanding assignment should replace the inquiry tasks from FILE, which was used in earlier validation studies (Hulshof et al., 2005). Similar to the inquiry task, the scientific text understanding assignment tests inquiry skills with active research processes (e.g. selecting variables perform experiments). Due to the similarities between the inquiry task and the scientific text understanding assignment, it is expected that the assignments scores and the IST scores correlate. IST scores should have predictive value for the level of text understanding. Such a correlation would support findings of earlier studies that the IST scores and the inquiry tasks scores correlate. This would be additional support for the convergent validity of the IST.

*Summary: research question and hypothesis*

The goal of the study was to find support for the validity of the IST. The research question was: In how far is the IST a valid instrument to measure inquiry skills? Therefore, it was hypothesized that IST scores would correlate with science motivation and with understanding of a scientific text. A moderate to high correlation between scores on the IST and the level of text understanding and between the IST scores and the DSMQ scores would support the interpretation of the IST score.

## Method

### *Respondents*

In total, 44 eleventh-grade pre-university students (5 VWO) participated in the study. Twelve of them were female, 29 were male. Three learners did not report their gender. The mean age of the learners was 17 (SD = 0.46). One learner missed the first test session due to sickness. In total, eight learners were deleted listwise, because they did not fill in all three tests; one only completed the biographic questions; three did only fill in the scientific text understanding assignment and six learners only completed the IST and the DSMQ.

Two secondary schools from the eastern part of the Netherlands volunteered to participate with their Technasium classes (henceforth referred to as School A and School B). The Technasium class is a very practice oriented educational branch (Technasium, n.d.). It can be chosen on the level of HAVO (higher general secondary education) and VWO (pre-university education). The learners get insight in scientific inquiry. In groups, they perform research projects, which they plan and implement with some guidance of teachers and research experts. With the help of research experts from universities and companies they also write scientific research reports on these projects. These learners were chosen, because they have more experience with science than the average secondary education student. This was regarded as an advantage with regard to the scientific text understanding assignment, which requires knowledge on basic scientific terms (e.g. variables).

The class from School A contained 16 learners and the class from school B contained 28 learners. All learners agreed to participate in the study and all received permission from their parents. As a service in return, both schools received feedback on the test results of their Technasium classes. Additionally, each learner got individual feedback on his or her performance. Also, the research assistant gave a lesson about quality criteria for research to the class from School A.

### *Instruments*

*Inquiry Skills Test.* In this study, the revised version of the IST was used, because it has a higher reliability than the original version by Horstink (von Ruedorffer, 2007). The IST contains two sets of subtests: (a) the WGKDT related subtests, namely *Conclusion* and *Interpretation* and (b) the TIPSII related subtests *Identify Variables*, *Stating Hypothesis*, *Operational Definitions*, *Designing Investigations* and *Graphing and Interpreting Data*.

The subtest *Conclusion* contains three expositions. Learners had to rate five to six statements about each exposition. They have to choose between five response possibilities. A



statement can be “true”, “probably true”, “probably false”, “false” or “not enough information”. Figure 1 displays an example of the subtest conclusion.

The subtest *Interpretation* contains six expositions and interpretations of these expositions. An exposition has to be regarded as true and with regard to this the learners have to determine whether an interpretation is “true” or “false”. In Figure 2 an example of this subtest can be seen.

The third subtest of the IST is the TIPSII related part. It contains 52 multiple choice questions with four answer possibilities. The questions of the five subtests *Identify Variables*, *Stating Hypothesis*, *Operational Definitions*, *Designing Investigations* and *Graphing and Interpreting Data* are mixed randomly. There is always only one correct answer. Some of the questions contain graphics or data that have to be interpreted by the learners. An example of this subtest is presented in Figure 3.

The IST has no time restrictions. It takes about 55 minutes to fill in the extended version of the IST (von Ruedorffer, Streese, Kamps & Schmitt, 2007). It was transferred to an online questionnaire program (van Rixtel, 2010).

*Uiteenzetting:*

Een lerares Engels bekeek met de leerlingen uit één van haar klassen de film die gemaakt werd naar Charles Dickens’ boek “Great Expectations”, terwijl de leerlingen uit al haar andere klassen alleen het boek bestudeerden, zonder de film te zien. Ze wilde weten of films effectief gebruikt konden worden in het literatuuronderwijs. Direct na iedere les werden aan deze twee groepen door middel van een toets de waardering (D: Bewertung) van en het inzicht (D: Erkenntnis) in het verhaal vastgesteld. Op beide toetsen scoorde de klas die de film had gezien hoger. Deze klas raakte zo

| <i>Mogelijke conclusies:</i>  | <b>W</b> | <b>WW</b> | <b>OI</b> | <b>WO</b> | <b>O</b> |
|---|----------|-----------|-----------|-----------|----------|
| <b>1.</b> De toetsen die tijdens dit experiment werden afgenomen, waren bedoeld om meer dan alleen de feitenkennis (D: Faktenwissen) over het boek te beoordelen. |          |           |           |           |          |
| <b>2.</b> De leerlingen die les kregen met behulp van de film, kregen de opdracht aan het begin van het schooljaar het boek te lezen.                             |          |           |           |           |          |

Figure 1. An exposition with two conclusions from the subtest Conclusion from the WGKDT.

***Uiteenzetting:***

Een verkoper van Dermatrix Lotion verkondigde dat zijn product in een handomdraai spierpijn zou verlichten, door in de pijnlijke lichaamsdelen door te dringen. De verkoper bracht tien druppels lotion op een dik stuk schoenleer aan, dat de lotion al snel opnam.

| <i>Mogelijke interpretaties:</i>  | <b>Juist</b> | <b>Onjuist</b> |
|---|--------------|----------------|
| <b>1.</b> De verkoper toonde de genezende werking van het product aan.  |              |                |
| <b>2.</b> Het was de bedoeling van de verkoper te suggereren dat als de lotion door een dik stuk schoenleer heen kon dringen, die ook door zou kunnen dringen in pijnlijke spieren. |              |                |
| <b>3.</b> De demonstratie van de verkoper was een bewijs voor zijn bewering dat de lotion spierpijn kan verlichten.   |              |                |

*Figure 2.* An exposition with three interpretation from the subtest Interpretation of the WGKDT.

1. Jim denkt dat zijn basketbal hoger zal stuiteren als de luchtdruk in de bal toeneemt. Om deze hypothese te toetsen verzamelt hij een aantal basketballen en een luchtpomp met een drukmeter. Hoe kan Jim zijn hypothese testen?
- A. Door basketballen met verschillende kracht vanaf dezelfde hoogte te stuiteren.
  - B. Door basketballen met verschillende luchtdruk vanaf dezelfde hoogte te stuiteren.
  - C. Door basketballen met dezelfde luchtdruk onder verschillende hoeken ten opzichte van de vloer te stuiteren.
  - D. Door basketballen met dezelfde luchtdruk vanaf verschillende hoogtes te stuiteren.

*Figure 3.* One of the TIPSII related items with four answer possibilities.

*Dutch Science Motivation Questionnaire.* The DSMQ is a questionnaire for science motivation (Wilhelm et al., submitted). It contains 30 Likert-Scale items with four options (i.e., totally agree, agree, disagree, totally disagree). The learners had to answer the question

with the option that fits their opinion the best. An example of the DSMQ is displayed in Figure 4. It takes about 5 to 10 minutes to complete the DSMQ (Wilhelm et al., submitted). The DSMQ was also transferred to the online questionnaire program (van Rixtel, 2010).

|   |               |          |                   |
|---|---------------|----------|-------------------|
| 2. Mijn leraren helpen mij om zelf te kunnen beoordelen hoe ik onderzoek doe. |               |          |                   |
| Helemaal niet mee eens  | Niet mee eens | Mee eens | Helemaal mee eens |

Figure 4. A question from the DSMQ with the four answer possibilities.

*Scientific text understanding assignment.* This task was developed by the researcher and based on the IST and its subscales for the sake of this study. The complete task can be found in Appendix A. First, a text was chosen. The scientific text was based on a master thesis from the University of Twente about the influence of music and wall color on eating behavior (van Zoelen, 2010). This neutral topic was chosen on purpose to prevent any negative impact on performance of an emotional topic like bullying. An existing study was used to make the results more realistic. Some parts (e.g. about the influence of music speed on eating behavior) were deleted to create a text that would be no longer than ten pages. The questionnaire was developed, based on the subtests of the IST. A table was developed listing all subtests of the IST. For each subtest, possible open questions were developed; for instance, the question belonging to the subtest *Identify Variables* was “Which variables should be included in the study to answer the research question?” (see Appendix B). The goal was to create a questionnaire with three questions for each subscale. Based on these questions, adaptations to the text were made. Paragraphs were rewritten, “mistakes” were included and the text and results were simplified to adapt it to the competence level of 16 year-olds. Possible unknown words were defined. For the open questions, possible answers were formulated. The questions were also given to a research expert. Based on his feedback, some questions were deleted and some new questions were designed. This process was repeated several times. Each time, questions and text were adjusted to yield a short, clear and condensed text from which all inconsistencies were dried out. Additionally, the layout of the questionnaire was adjusted to make it more structured and an instruction text was added to give instructions about how to perform the assignment.

The final product was tested with two students in a pilot study. Both were female and had a Bachelor’s degree in Psychology. Based on the results and the feedback of the first

student two questions were adjusted, because they were ambiguous. Moreover, spelling mistakes in the text were corrected. The new version was given to the second student who also filled in the questionnaire and provided feedback. This time only the layout was adjusted to improve the structure of the questionnaire. The second student perceived none of the questions as ambiguous. It took both pilot-testers about 45 minutes to fill in the task.

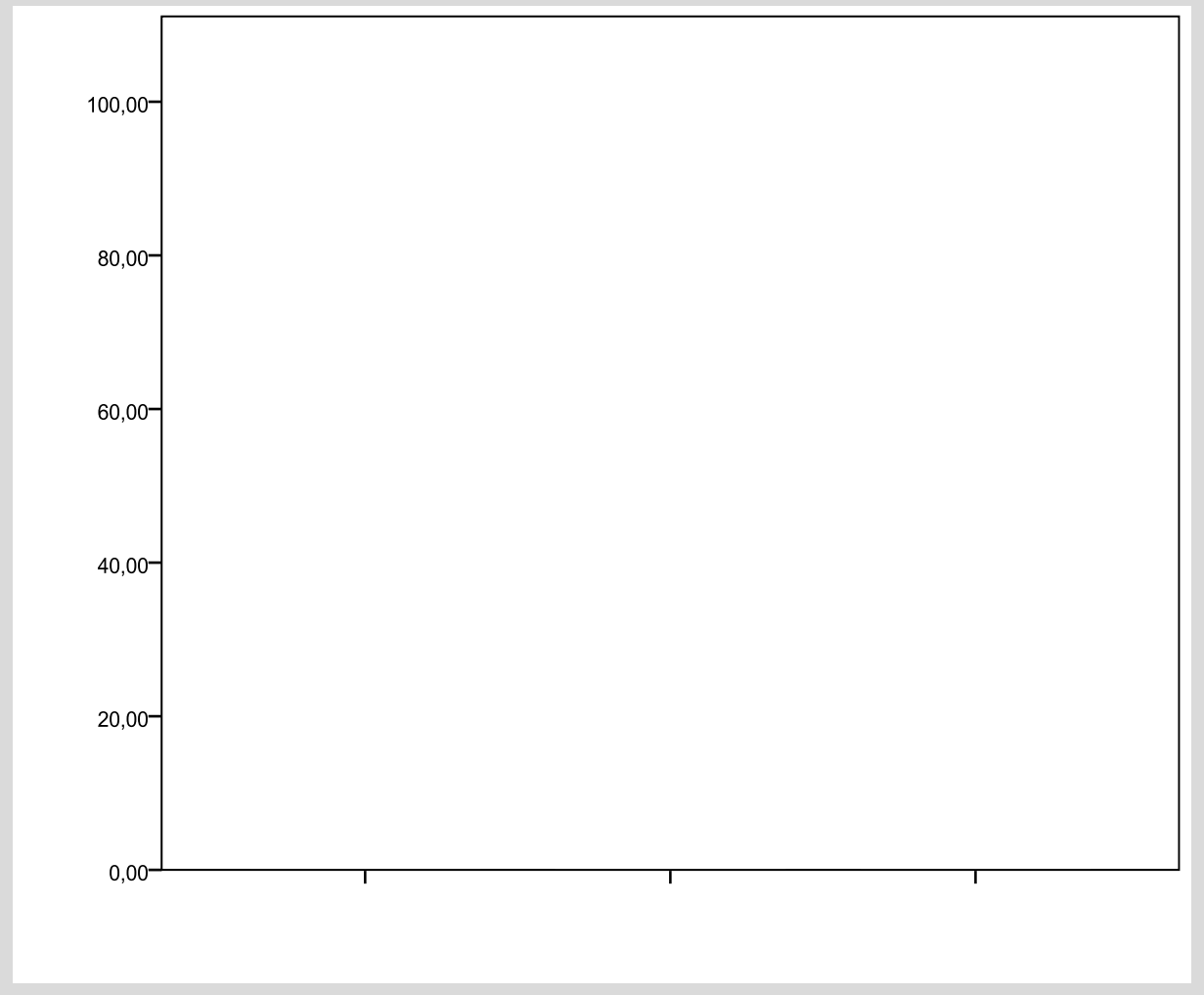
In the text understanding assignment learners receive an edited scientific paper of ten pages accompanied by 17 open questions. In the article, a study about the influence of music on eating behavior is described. It is divided in four parts, namely *Introduction*, *Hypothesis*, *Method*, *Results* and *Conclusion and Discussion*.

Two to six questions have to be answered with regard to each part. Some questions ask the learners to find specific information about the study in the text, like question 1.1: “Below, name all variables from the different studies that are discussed in the introduction.” In other questions the learners have to find mistakes in the design of the study. One item asks the learners to draw a graph (Figure 5). The questions have to be answered in a predetermined order. The learners are not allowed to look back to previous questions. To prevent this, the answers on each part (i.e. *Introduction*, *Hypothesis*, *Method*, *Results* and *Conclusion and Interpretation*) are collected by the research assistant after they are finished. In the *Result* and the *Conclusion and Discussion* parts of the article answers on the question from the *Introduction*, the *Hypothesis* and the *Method* can be found. Therefore the article is divided into two parts. Learners had to raise their finger when they finished the first three subtests. Then they received the second part of the text. The pilot tests revealed that it takes about 45 minutes to complete the text understanding assignment.

### *Procedure*

The test was divided into an online part and a paper-pencil part. The online part included the IST and the DSMQ whereas the scientific text understanding assignment was presented in a printed version. At School A, the tests were taken in two sessions during school time. The online part took place in a computer room and the scientific text understanding assignment took place five days later in a regular class room. 60 minutes were appointed for the online part and 90 minutes for the evaluation test. At School B, all tests were taken at one afternoon after school time in two computer rooms. The learners took no break between the online version and the evaluation test.

4.3 Stel dat Hypothese 3 (H3: Het muziektempo beïnvloed de hoeveelheid gezond voedsel die geconsumeerd wordt) ook onderzocht werd en de resultaten waren als volgt: In de snelle muziek conditie wordt significant meer gezond voedsel geconsumeerd dan in de controle conditie en de langzame muziek conditie. Er werd geen significant verschil tussen de controle conditie en de langzame muziek conditie gevonden. Schets hieronder hoe een grafiek uit zou kunnen zien. Gebruik de grafiek uit de tekst als voorbeeld.



*Figure 5. Open question from the Graphing and Data Interpretation subtest of the scientific text understanding assignment.*

At the beginning of the online part, the learners received a short introduction about the goal of the study and about the procedure. All learners received links to the website where they could find the IST and the DSMQ. The research assistant walked around to help learners who had problems to open the website. She was available during the whole test situation to answer questions about the procedure and to make sure the learners filled in the tests individually. The research assistant did not answer questions about the content of the tests. At

both schools, a teacher was present at the start and the end of the study to remember the learners to remain calm and follow the orders. After 45 minutes, the learners that had not yet finished the IST and the DSMQ were asked how far they had got. In case learners answered only about half of the items after 45 minutes, the research assistant motivated the learners to complete the rest of the questionnaires. After 90 minutes, all learners had completed the online part. After another 90 minutes all learners also completed the written part of the test.

### *Scoring*

In the first step, the data from the online questionnaire and from the paper-pencil part were brought together in a SPSS file. Individual codes containing the last letters from sir name, name and the date of birth were used to connect test scores. The IST data was transformed as in Horstinks' study; for correct answers one point was given and an incorrect answer was credited with zero points. Afterwards the points were computed to subtest scores. The subtest scores were then summed up to a total score of the IST. IST scores ranged from zero to 84 points. For the DSMQ total score, the mean score for each learner was calculated (Wilhelm et al., submitted). The mean scores could range from 1.00 to 4.00. No missing values were possible, because on the online questionnaire software learners had to answer all questions before they could reach the next test page.

An evaluation scheme was used to analyze the open questions of the scientific text understanding assignment (see Appendix C). The development of the evaluation criteria followed a stepwise procedure. It was based on the answers the learners gave to the questions. For each question on the scientific text understanding assignment, a learner could gain between zero and three points. Whether a learner receives zero, one, two or three points depended on how close the learner's answer resembled the sample solution. The evaluation scheme contained detailed instructions and examples about how to rate different answers. Subtracting points for incorrect parts of answers led to the effect that many learners received zero points. Therefore it was decided not to subtract points for incorrect parts of answers. It was not expected that this would threaten the validity of the results, because zero points were given when no answer was given or for completely incorrect answers, one point was given for partly correct answers, two points were given for mostly correct answers and three points were given for (nearly) correct answers. How many points a learner received was therefore dependent on his level of correct text understanding. For the total score, the scores on all sub questions were summed up. On the scientific text understanding assignment a total score between zero and 51 points was possible. To test the rating scheme for the scientific text

understanding assignment, the inter-rater reliability was determined. A second rater evaluated the answers of six learners (13.63%). The results were compared to those of rater one. In total 102 pairs of scores were compared. The Cohen's kappa was calculated with a crosstab analysis. The consensus between the two raters was .96.

### *Data analysis*

First, outliers were deleted if necessary. In general, outliers are scores that are more than two or three standard deviations higher or lower than the mean (Osborne & Overbay, 2004). Osborne and Overbay state that it cannot simply be assumed that outliers are caused by fraud or by mistakes. Therefore, each outlier has to be checked separately. Additionally it was tested whether the sample distributions were normal. After that, the descriptive statistics were determined. To test internal consistency, the Cronbach's alpha of the tests and subtests were determined and compared to earlier studies.

Correlations between scores on the three tests were calculated. As a parametric test Pearson correlation was used and as a non-parametric measure Spearman's rho and Kendall's tau\_b were calculated. Parametric tests are used for data that is normally distributed and non-parametric tests are used for data that is not normally distributed. Parametric tests are more accurate. Bigger samples are approximately normally distributed. This is true for samples with  $n > 30$ . In this study, for a normal distributed sample the Pearson correlation was used. For samples that were not normal distributed non-parametric measures and parametric measures were both used and the results were compared, because the sample in this is with  $n > 30$  expected to be approximately normal distributed.

Significance was tested one-sided because a positive correlation between the IST and the DSMQ on the one hand and the IST and the text understanding assignment on the other hand was expected. A multiple linear regression analysis was performed to clarify how much of the variance in IST scores was explained by scores on the other tests.

## Results

### *Descriptive Statistics*

First, the outliers were deleted listwise. In this study, five cases had scores more than two times the standard deviation lower than the mean on at least one of the tests. Two cases had scores more than three times the standard deviation lower than the mean on the IST and the DSMQ. A closer look at the data made clear that these learners selected the same answer on each question. This was not the case for the other three learners. Therefore two outliers were deleted. The Shapiro-Wilk Test was used to test how the scores of the remaining 34 learners on the tests were distributed. The scores on the scientific text understanding assignment had a normal distribution ( $Z = .97$ ,  $p = .45$ ). The distribution of the IST scores, however differed significantly from an expected normal distribution ( $Z = .86$ ,  $p < .01$ ). The same applied to the DSMQ scores ( $Z = .90$ ,  $p < .01$ ).

Two age statements were implausible. One learner stated to be over 100 years old, whereas the second student claimed to have a negative age. The age statements of these two learners were deleted pairwise, because the age statements were not essential to test the hypotheses. The mean age of the 32 valid cases was 16.97 ( $SD = 0.65$ ). Twenty-two of the learners were male and 12 were female. 15 learners from School A and 19 learners from School B remained. The mean score on the IST was 61.44 ( $SD = 12.16$ ). On the DSMQ the mean score was 2.97 ( $SD = 0.38$ ) and on the text understanding assignment 22.65 ( $SD = 7.74$ ). Table 1 displays the descriptive statistics.

Table 1

*Descriptives IST, DSMQ, and Scientific Text Understanding Assignment*

| <b>Subtest</b>                           | <b>N</b> | <b>Maximum Possible Points</b> | <b>MEAN</b> | <b>SD</b> | <b>Lowest Score</b> | <b>Highest Score</b> |
|--|----------|--------------------------------|-------------|-----------|---------------------|----------------------|
| IST                                      | 34       | 84                             | 61.44       | 12.16     | 26                  | 77                   |
| TIPSII                                   | 34       | 52                             | 39.94       | 9.88      | 13                  | 51                   |
| WGKDT                                    | 34       | 32                             | 21.50       | 3.22      | 13                  | 27                   |
| DSMQ                                     | 34       | 3                              | 2.97        | 0.38      | 1.63                | 4                    |
| Scientific Text Understanding Assignment | 34       | 51                             | 22.65       | 7.74      | 7                   | 37                   |

### *Group differences*

It was tested if there were significant differences between schools and gender. An ANOVA with school as the independent variable showed that there was a significant school difference



on the IST ( $F = 5.65$ ,  $p < .05$ ). The learners from School A scored significantly higher ( $M = 66.67$ ,  $SD = 5.22$ ) than the learners from School B ( $M = 57.32$ ,  $SD = 14.47$ ). On the text understanding assignment the learners from School A ( $M = 27.93$ ,  $SD = 4.56$ ) also scored higher than the learners from School B ( $M = 18.47$ ,  $SD = 7.21$ ,  $F = 19.58$   $p < .01$ ). With regard to the text understanding scores, there was a significant gender difference ( $F = 7.25$ ,  $p < .05$ ). The women ( $M = 27.08$ ,  $SD = 4.38$ ) reached significantly higher scores than the men ( $M = 20.23$ ,  $SD = 8.16$ ). No significant gender differences in the IST scores were found. Moreover, there were no significant differences for school or for gender in the DSMQ scores. It was decided to test the sample as a whole group and not to separate it by schools. Firstly, a separated sample would be very small. Secondly, there were no differences in DSMQ scores, therefore it was not expected that overall school differences would influence the results.

### *Internal Consistency*

Tabel 2 displays the Cronbach's alpha coefficients of the tests. SPSS automatically uses the Kuder-Richardson Formula for dichotomous items for the IST scores. The IST as well as the DSMQ had a high Cronbach's alpha ( $\alpha = .91$  and  $\alpha = .91$ , respectively). The subtests of the IST that belong to the TIPSII related part also had, with one exception, a high reliability, ranging from .74 to .82. Only the subtest *Graphing and Interpreting Data* had a moderate Cronbach's alpha ( $\alpha = .50$ ). The alphas of the subtests that belonged to the WGKDT were, however, relatively low ( $\alpha < .05$ ). The subtest *Conclusion* also had a low reliability ( $\alpha = .49$ ) and the subtest *Interpretation* even had a negative coefficient ( $\alpha = -.02$ ). The Cronbach's alpha of the scientific text understanding assignment was relatively high ( $\alpha = .79$ ). By deleting on item (1.2) it could at maximum be increased to 0.80 ( $\alpha = .80$ ). This small improvement did not justify a reduction of the number of items.

### *Correlational analysis on total scores*

Correlation analysis was done for the IST, the DSMQ and the scientific text understanding assignment. The analysis of results was one-sided, because a positive correlation was expected. A significance level of  $\alpha = .05$  was handled. Table 3 displays the results. Parametric and non-parametric tests revealed similar results. No significant correlation was found between the IST and the DSMQ ( $r = .22$ ,  $p = .11$ ). The text understanding assignment correlated significantly with the IST ( $r = .50$ ,  $p < .01$ ) and with the DSMQ ( $r = .38$ ,  $p < .05$ ). There was only a slight difference in non-parametric test results. Here, the scientific text understanding assignment correlated stronger with the DSMQ than with the IST.

Table 2

*Cronbach's Alpha of the IST, the DSMQ, the Scientific Text Understanding Assignment and Their Subtests*

| <b>Subtest</b>                           | <b>Cronbach's Alpha</b> | <b>N of Items</b> |
|--|-------------------------|-------------------|
| IST                                      | .91                     | 84                |
| TIPSII                                   | .93                     | 52                |
| <i>Identify Variables</i>                | .82                     | 12                |
| <i>Stating Hypothesis</i>                | .79                     | 10                |
| <i>Operational Definitions</i>           | .74                     | 10                |
| <i>Designing Investigations</i>          | .82                     | 10                |
| <i>Graphing and Interpreting Data</i>    | .50                     | 10                |
| WGKDT                                    | .42                     | 32                |
| <i>Conclusion</i>                        | .49                     | 16                |
| <i>Interpretation</i>                    | -.02                    | 16                |
| DSMQ                                     | .91                     | 30                |
| Scientific Text Understanding Assignment | .79                     | 17                |

Table 3

*Parametric and Non-parametric Correlation Analysis for the IST, the DSMQ and the Scientific Text Understanding Assignment*

|                                   | <b>Subtest</b>                           | <b>IST</b> | <b>DSMQ</b> |
|-----------------------------------|--|------------|-------------|
| Kendall's Tau_b<br>(1-tailed)     | DSMQ                                     | .33        |             |
|                                   | Scientific Text Understanding Assignment | .29**      | .33**       |
|                                   |  |            |             |
| Spearman's Rho<br>(1-tailed)      | DSMQ                                     | .32        |             |
|                                   | Scientific Text Understanding Assignment | .39*       | .50**       |
|                                   |  |            |             |
| Pearson Correlation<br>(1-tailed) | DSMQ                                     | .22        |             |
|                                   | Scientific Text Understanding Assignment | .50**      | .38*        |
|                                   |  |            |             |

\*  $p < .05$ . \*\*  $p < .01$ .

#### *Subscale correlation between IST and the scientific text understanding assignment*

The scientific text understanding assignment is based on the IST. For each subtest of the IST, corresponding items for the scientific text understanding assignment were developed. Only *Conclusion* and *Interpretation* were joined together in one subtest. Therefore, the correlation between the scores on the subtests of the IST and the related items of scientific text understanding assignment were analyzed to see if they measure the same constructs. The analysis of results was two-sided, because no hypothesis about the direction of the correlation

was made. The results are displayed in Table 4. There were significant positive correlations between the scores of the TIPSII related subparts of the IST and the TIPSII related subparts of the scientific text understanding assignment ( $r = .41, p < .05$ ), including the subtests *Identify Variables*, *Stating Hypothesis*, *Operational Definitions*, *Designing Investigations* and *Graphing and Interpreting Data* of the IST and the understanding assignment. Moreover, there were significant correlations between the scores on the subtests *Identify Variables* ( $r = .41, p < .05$ ), *Stating Hypothesis* ( $r = .51, p < .01$ ), *Designing Investigations* ( $r = .39, p < .05$ ) and *Graphing and Interpreting Data* ( $r = .44, p < .01$ ). The correlation between the subscale scores of the scale *Operational Definitions* ( $r = .24, p = .17$ ) was not significant. Since the subtests *Conclusion* and *Interpretation* of the IST were joined together in one subtest on the text understanding assignment, it was not reasonable to test the correlation for these subtests apart. Therefore, the correlation with the total score was tested. No significant correlation was found ( $r = .22, p = .22$ ).

Table 4

*Parametric Correlation Analysis between the subscales of the IST and the scientific text understanding assignment*

| <b>IST</b>                            | <b>Scientific text understanding assignment</b> | <b>Pearson Correlation (2-tailed)</b> |
|---------------------------------------|---|---------------------------------------|
| Total                                 | Total   | .41*                                  |
| <i>Identify variables</i>             | <i>Identify variables</i>                       | .41*                                  |
| <i>Stating hypothesis</i>             | <i>Stating hypothesis</i>                       | .51**                                 |
| <i>Operational definitions</i>        | <i>Operational definitions</i>                  | .24                                   |
| <i>Designing investigations</i>       | <i>Designing investigations</i>                 | .39*                                  |
| <i>Graphing and interpreting data</i> | <i>Graphing and interpreting data</i>           | .44**                                 |
| Conclusion and Interpretation         | Conclusion and Interpretation                   | .22                                   |

\*  $p < .05$ . \*\*  $p < .01$ .

Table 5

*Parametric Correlation Analysis Between the IST Total Scores and the DSMQ Sub Constructs*

| <b>DSMQ</b>                      | <b>IST</b> |
|----------------------------------|------------|
| <i>Intrinsic task motivation</i> | .28        |
| <i>Utility value</i>             | .32*       |
| <i>Competence belief</i>         | .06        |
| <i>Feedback</i>                  | .10        |
| <i>Self-efficacy</i>             | .27        |
| <i>Goal setting</i>              | .17        |
| <i>Peer support</i>              | -.11       |
| <i>Performance motivation</i>    | .19        |

\*  $p < .05$ . \*\*  $p < .01$ .

*Correlation between IST total score and DSMQ sub constructs*

The correlations between the IST total score and the DSMQ sub constructs were examined, to determine whether the IST scores correlate with any aspect of science motivation. The analysis of results was two-sided, because no hypothesis about the outcome was formulated before. The results are displayed in Table 5. Only the sub construct *utility value* correlated significantly with the IST ( $r = .32, p < .05$ ).

*Multiple linear regression analysis*

For the regression analysis the ‘stepwise’-method was used. The IST scores as well as the DSMQ scores correlated with the text understanding scores. However the scores of the IST and the DSMQ did not correlate. Therefore, the scientific text understanding assignment was chosen as the dependent variable and the IST and the DSMQ were chosen as the independent variables. This analysis revealed that the IST explained 25.3% of the total variance in the scientific text understanding assignment and the DSMQ explains additional 7.7%. In total, the IST and the DSMQ explained 33.0% of the variance in the scientific text understanding assignment. The contribution of the IST to the explained variance was significant ( $\beta = .44, p < .01$ ), whereas the contribution of the DSMQ to the explained variance was not significant ( $\beta = .29, p = .07$ ).

## Conclusions and Discussion

Goal of the study was to collect evidence to support the validity of the IST. The IST and two additional tests were administered to a sample of pre-university students. It was hypothesized that IST scores would correlate positively with scores on a scientific text understanding assignment and with DSMQ scores. Such results would support the interpretation of the IST score.

As expected, IST scores were related to scores on the scientific text understanding assignment. This correlation supported the validity of the IST. The same relation as between the IST and the scientific text understanding assignment was found for their subscales. The scores on four out of the six subtests of the IST (*Identify Variables*, *Stating Hypothesis*, *Designing Investigations* and *Graphing and Interpreting Data*) were related to the scores of their corresponding subscales on the IST. No significant correlations were found between the IST subscales *Operational Definitions* and *Conclusion and Interpretation* and their corresponding subtests of the scientific text understanding assignment. The hypothesis that inquiry skills and research motivation were related could not be confirmed. Contrary to the expectations, there was no support that the IST and the DSMQ measure related constructs. Only the sub construct *utility value* seems to be related to the inquiry skills measured by the IST, because the total scores of the IST correlated with this sub construct of the DSMQ. Also, DSMQ scores were related to scores on the scientific text understanding assignment. Although the IST scores and the DSMQ scores did not correlate, IST scores as well as DSMQ scores correlated with scores on the text understanding assignment. Therefore a regression analysis was performed with the IST scores and the DSMQ scores as the independent variables and the text understanding scores as the dependent variable. The IST explained 25.3% of variance in scientific text understanding assignment. The DSMQ explained an additional 7.7% of the variance in the scientific text understanding assignment. Together, the IST and the DSMQ explain 33% of the variance in the scores. It was however remarkable, that the DSMQ scores do not contribute significantly to the explained variance in the assignment scores. With regard to this, the IST wins above the DSMQ. The results of the regression analysis are a support for the discriminate validity of the IST. The regression analysis supports the hypothesis, that inquiry skills as measured in the IST also play a role in scientific text understanding.

The results of the study are in line with the results of earlier studies (Horstink, 2005; von Ruedorffer et al., 2007). In these studies, correlations between IST scores and scores on an inquiry task were found. It was hypothesized that the IST as well as the inquiry task

demand similar skills. In this study the inquiry task was replaced by a different task, which was also expected to call upon inquiry skills. Therefore, the correlation between the IST scores and the assignment scores support findings of earlier studies. Horstink et al. and von Rueddorffer et al. also found that the IST contributed significantly to the explanation of the variance in scores of the inquiry learning task. Similar to that, the IST is a significant predictor of scores on the text understanding assignment.

Although Lange et al. (2010), did not find a correlation between IST scores and DSMQ scores in an earlier study, such a correlation was hypothesized due to limitations of the study of Lange et al. and due to other research on the topic (Glynn & Kabolla, 2007; Zusho, Pintrich & Coppola, 2003; Faccione, 2000). However, the results of this study confirm the findings of Lange et al. Moreover, other validation studies that tested the relation between the IST and interest in science also found no correlation (Kip, Looge, Fens & Heilema, 2009). Though, it is remarkable that there was no correlation between the IST scores and the DSMQ scores, whereas the DSMQ scores and the scores on the understanding assignment did correlate. With regard to this, it is questionable if it was reasonable to assume that science motivation and inquiry skills would be related in the first place. As mentioned studies on science success and motivation in general do show that motivation influences success on abstract and domain dependent science projects and tasks (Glynn & Kabolla, 2007; Zusho, Pintrich & Coppola, 2003; Faccione, 2000). The correlation between science motivation and the scores on the scientific text understanding task corroborates these findings. The results of this study seem to indicate that the IST measures skills in a way science motivation does not play a role. The IST scores correlated with scores on only one sub construct of the DSMQ, namely *utility value*. According to Glynn and Koballa (2007), this sub construct can be compared with intrinsic and extrinsic motivation. Possibly, intrinsic and extrinsic motivation play a role in tasks like the IST. Science motivation might only be important to engage in active research. Zusho et al. (2003) found a correlation between motivation and success on science projects. In this regard intrinsic and extrinsic motivation might play a role in multiple choice tasks like the IST whereas the other aspects of science motivation that are measured in the DSMQ might only play a role in integrated and active tasks like the scientific text understanding assignment. Scholz and Zuell (2012) found that the non-response rate especially in open-ended questions is related to interest in a subject. The motivation to fill in questions correlates with the interest in the topic. Such a clear correlation with multiple choice questions was not found. This indicates that motivation plays a bigger role with regard to open-ended questions.

Another explanation for the missing correlation between IST and DSMQ scores could be the sample of Technasium learners, who are expected to have a general interest in science. Their scores on science motivation may be overall high and independent from their actual science skills. Compared to the findings of Wilhelm et al. (2010), the DSMQ scores seemed higher in this study. For learners in secondary education, the mean was 2.5 (SD = 0.44) and for university students it was 2.7 (SD = 0.40). Contrary in this study the mean was 2.9 (SD = 0.38). Yet, the standard deviations were similar, so this would not explain the correlation between the DSMQ scores and the scores on the understanding assignment.

The study had some strong points and some weak points that have to be taken in to account. In comparison to earlier studies, the sample size (n= 44) was sufficient. The data of ten of the participants had to be deleted, because they did not fill in all scales. However, the analyses were done with a sample of 34 learners, which is still sufficient. There were no missing values, because the learners had to answer all items in the online test. This may have led to the phenomenon that learners became frustrated and start to select answers at random. This would have led to unreliable results, which would have influenced the results of the study. All students needed more than half an hour to fill in the online tests. Therefore, it is expected that they fill in the test seriously. Moreover, outliers were deleted, to reduce such effects. Also, the study reveals clear results. It is not reasonable that such results are caused by random selection of answers.

Another limitation was the testing situation at one of the schools. The learners had to fill in all tests without a break on their last day before summer vacations. Furthermore, the test took place after school during their free time. Some of the learners turned very frustrated and had to be motivated to complete the study. Additionally, the group was distributed over two rooms. Therefore one part of the group at a time was without supervision of the research assistant. The learners may not have filled in the questionnaires individually. Yet, the research assistant switched between the rooms. It can be assumed that most of the time the learners of both groups worked individually.

A strong point of the study was the high reliability of the instruments used. The analysis of the results shows that the IST has a high reliability with a Cronbach's alpha of .91. The DSMQ also had a high Cronbach's alpha of .91. Although the scientific text understanding assignment was newly developed, it had a high Cronbach's alpha (.79). The Cronbach's alpha of the IST is in line with earlier studies on the IST (Von Rueddorfer et al., 2007). However, with regard to the IST subtests *Conclusion* and *Interpretation* Cronbach's alphas were lower than .05. This also resembles results from earlier studies (e.g. Horstink,

2005; Geerding, Rijken & Vennemann, 2009). It is therefore reasonable to recommend that these subtests should not remain in the test. Contrary to the TIPSII related parts, the WGKDT related parts have a low reliability and therefore do not contribute to the validity of the test. The adaptations von Rueddorfer et al. (2007) made could not improve this effect. Given a large enough sample size, a factor analysis would detect to what extent the subscales of the IST measure different aspects of inquiry skills. Moreover, the subtests *Conclusion* and *Interpretation* from the IST were joined to one test (*Conclusion and Interpretation*) on the scientific text understanding assignment. In the assignment, the subtest included three questions. Possibly these questions did not clearly map the skills of one of the subtests *Conclusion* or *Interpretation*. It would have been better to formulate two to three questions corresponding with each subtest. As mentioned above, the total score of the IST subtests *Conclusion* and *Interpretation* did not correlate with the scores on the assignment subtest *Conclusion and Interpretation*. Therefore, it is time to replace them in the IST.

Summarized, the results of the study support the interpretation of the IST scores. The convergent validity of the IST is supported by the correlation between the IST scores and the scores on the scientific text understanding assignment. Although no correlation between the IST scores and the DSMQ scores was found this not by all means evidence against the validity of the IST, because it may be that such a correlation, was not to be expected to exist in the first place. As mentioned above, it may be that science motivation plays a role in active science tasks like the scientific text understanding assignment and not in tasks like the IST. Although the IST scores as well as the DSMQ scores both correlate with the text understanding assignment scores, it turned out, that only the IST significantly predicts the scores on the text understanding assignment. This supports the discriminant validity of the IST. It was expected that the IST measures skills that also play a role in the text understanding assignment. The correlation between IST scores and assignment scores supports that. However, the assignment scores also correlate with science motivation. Therefore, the DSMQ also measures skills that play a role in the text understanding assignment. That inquiry skills better predict scores on the text understanding assignment than science motivation, confirms the expectations. The IST seems to be a distinct instrument for inquiry skills, that could be used in school settings to help teachers detect differences in learners inquiry skills and to adapt their teaching strategies.

Further studies should again relate IST scores with performance on the scientific text understanding assignment to replicate the results. This should be done with different age groups and also with learners in pre-university education who do not participate in the



Technasium branch. Through this, the results could be generalized to other settings. Finally, a factor analysis is highly recommended to further validate the subscales of the IST, especially the subscales *Conclusion* and *Interpretation*. This study made clear, that it is time to replace them with subscales that improve the validity of the IST.

## References

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Bonnstetter, R. J. (1998, September). Inquiry: Learning from the past with an eye on the future. *Electronic Journal of Science Education*, 3 (1). Retrieved 6 September, 2012, from <http://www.scholarlyexchange.org/ojs/index.php/EJSE/article/viewArticle/7595/5362>
- Borphy, J. E. (1988). On motivating students. In D. Berliner & B. Rosenshine (Eds.), *Talks to teachers* (pp. 201-245). New York, NY: Random House.
- Burns, J. C., Okey, J. R., & Wise, K. C. (1985). Development of an integrated process skill test: TIPSII. *Journal of Research in Science Teaching*, 22(2), 167-177.
- Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86, 175-218.
- De geschiedenis van het Technasium. (n.d.). Retrieved September 16, 2012, from <http://www.technasium.nl/Organisatie.aspx>
- de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 86(2), 179-201.
- Dewey, J. (1910). *How we think*. Boston, MA: D.C. Heath & Company.
- Dewey, J. (1933). *How we think. A restatement of the relation of reflective thinking to the educative process* (Rev. ed.). Boston, MA: D. C. Heath & Company.
- Dewey, J. (1938). *Logic: The theory of inquiry*. New York, NY: Holt, Rinehart and Winston.
- Dillashaw, F. G., & Okey, J. R. (1980). Test of integrated process skills for secondary science students. *Science Education*, 64, 601-608.

- Ennis, R. H., & Millman, J. (1985). *Cornell critical thinking test, level X*. Pacific Grove, CA: Critical Thinking Books and Software (formerly Midwest Publications).
- Evers, A. (1992). *Amsterdamse Bereopen Interessen Vragenlijst ABIV: Handleiding*. Lisse: Swets & Zeitlinger.
- Facione, P.A. (1990). *The California Critical Thinking Skills test and manual*. Millbrea: California Academic Press.
- Facione, P. A., & Facione, N. C. (1992). *The California Critical Thinking Disposition Inventory*. Millbrea, CA: California Academic Press.
- Facione, P. A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skills. *Informal Logic*, 20(1), 61-84.
- Fraser, B. J. (1979). *Test of enquiry skills*. Melbourne: Australian Council for Educational Research.
- Geerdink, J., Rijken, K., & Vennemann, C. (2009). *Validatie van Inquiry Skills Test. [Validity of the Inquiry Skills Test.]* Unpublished research manuscript, University of Twente.
- Glynn, S. M., & Koballa T. R., Jr. (2006). Motivation to learn in college science. In J. Mintzes & W. H. Leonard (Eds.). *Handbook of college science teaching* (pp. 25-32). Arlington, VA: National Science Teachers Association Press.
- Hensel, T., Kuipers, S., & Laseur, E. (2009). *Validatie Test Inquiry Skills*. Unpublished research manuscript, University of Twente.
- Horstink, M. (2005). *Constructie en validatie van een test voor het meten van inquiry skills. [Construction and validation of a test for the measurement of inquiry skills.]* Unpublished bachelorthesis, University of Twente.
- Hulshof, C. D., Wilhelm, P., Beishuizen, J. J., & van Rijn, H. (2005). FILE: a tool for the study of inquiry learning. *Computers in Human Behavior*, 21(6), 945-956.

- Kip, H., Looge, W., Fens, H., & Heilema, K. (2009). *Onderzoek naar de validiteit van de Inquiry Skills Test. [Research on the validity of the Inquiry Skills Test]*. Unpublished research manuscript, University of Twente.
- Kuhn, D. (1993). Connecting scientific and informal reasoning. *Merrill-Palmer Quarterly*, 39(1), 74-103.
- Lange, L., Von der Goltz, N., & Drawert, A. (2010). *Validatie test inquiry skills. [Validity test inquiry skills]*. Unpublished research manuscript, University of Twente.
- Lavinghousez, W. E. (1997). *Evaluation and assessment*. Retrieved 9 January, 2005, from: <http://pages.xtn.net/~blavingh/evaluation.html>
- Nederlandse organisatie voor wetenschappelijk onderzoek/programmaraad voor het onderwijsonderzoek. [Dutch organization for science research/programme for educational research.] (2003). *Programmaboek Onderwijsonderzoek 2004-2007*. Den Haag: NWO.
- Nederlandse Organisatie voor wetenschappelijk onderzoek/programmaraad voor het onderwijsonderzoek. [Dutch organization for science research/programme for educational research.] (2012). *Programmaboek Onderwijsonderzoek 2012-2015*. Den Haag: NWO.
- Okey, J. R., Wise, K. C., & Burns, J. C. (1982). *Integrated process skills test II*. University of Georgia, Department of science education.
- Osborne, J. W., & Overbay A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation*, 9(6). Retrieved October 21, 2012 from <http://pareonline.net/getvn.asp?v=9&n=6>
- Platform Bèta Techniek. (2008) *Technomonitor 2008*. Nijmegen: ResearchNed.

Reid, D. J., Zhang, J., & Chen, Q. (2003). Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning*, 19, 9-20.

Scalon, E., Anastopoulou, S., Kerawalla, L., & Mulholland, P. (2011). How technology resources can be used to represent personal inquiry and support students' understanding of it across contexts. *Journal of Computer Assisted Learning*, 27, 516-529.

Schaferman, S.D. (1991). *An introduction to critical thinking*. Retrieved, 27 february 2005, from <http://www.freeinquiry.com/critical-thinking.html>

Scholz, E., & Zuell, C. (2012). Item non-response in open-ended questions: Why does not answer on the meaning of left and right? *Social Science Research*, 41, 1415-1428.

Schönrock-Adema, J. (2002). *De ontwikkeling en evaluatie van een zelfinstructieprogramma voor een training in basisgespreksvaardigheden*. Groningen: University of Groningen.

Smith, P., & Whetton, C. (1992). *Critical reasoning tests*. Windsor: NFER-Nelson.

Wihlem, P., Stellmacher, O., Eysink, T. H. S., & ten Klooster, P. M. (submitted).  
Construction and validation of the Dutch Science Motivation Questionnaire.

Stokking, K. M., & van der Schaaf, M. F. (1999). *Beoordelen van onderzoeksvaardigheden van leerlingen: Richtlijnen, alternatieven en achtergronden*. Utrecht: Universiteit Utrecht, Onderwijskunde, ISOR.

Technasium (englich flyer). (n.d.). Retrieved September 16, 2012, from [http://www.technasium.nl/downloadfiles/technasium \(english flyer\).pdf](http://www.technasium.nl/downloadfiles/technasium%20(english%20flyer).pdf)

van Rixtel, J. (2010). *ThesisTools Online Enquêtes*. Retrieved November 7, 2012, from <http://www.thesistools.com/>

- von Ruëddorfer, B., Streese, E., Kamps, L., & Schmitt, K. (2007). *Validatie van de Inquiry Skills Test. [Validation of the Inquiry Skills Test.]* Unpublished research manuscript, University of Twente.
- van Zoelen, C. (2010). *Weet in welke omgeving je eet. De invloed van muziek en kleur op eetgedrag. [Know the environment you eat in. The influence of music and colour on eating behavior.]* Unpublished Masterthesis, University of Twente.
- Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction, 14*, 89-109.
- Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber & A. S. Carton (eds.). *The collected works of L. S. Vygotsky, Volume 1: Problems of general psychology* (Rev. ed.). New York, NY: Plenum.
- Vygotsky, L. S. (1981). The genesis of higher mental functions. In J. V. Wertsch (Ed.), *The concept of activity in Soviet Psychology* (pp. 144-188). Armonk, NY: Sharpe.
- Walma van der Molen, J. H. (2007). De ontwikkeling van een attitude-instrument op het gebied van wetenschap en techniek voor leerlingen in het basisonderwijs. Platform Bèta Techniek.
- Watson, G., & Glaser, E. M. (1964). Intelligentie en probleemaanpak. *De Psycholoog, 29*, 223-228.
- Watson, G., & Glaser, E. M. (1994). *Watson-Glaser Critical Thinking Appraisal-Manual Form S*. San Antonio, TX: Psychological Corporation.
- Wells, G. (1999). Dialogic inquiry in education: building on the legacy of Vygotsky. In C. D. Lee & P. Smagorinsky (Eds.), *Vygotskian perspectives on literacy research* (pp. 51-84). New York, NY: Cambridge University Press.

Wilhelm, P. (2001). *Knowledge, skills and strategies in self-directed inductive learning*.

Doctoral dissertation, Leiden University.

Zanten, D.L. van, Dekker, R., & Berkhout, C. (1997). *Watson-Glaser kritisch denken test*.

Lisse: Swets & Zeitlinger.

Zusho, A., Pintrich, P. R., & Coppola, B. (2003). Skill and will: The role of motivation and

cognition in the learning of college chemistry. *International Journal of Science*

*Education*. 25(9), 1081-1094.

**Appendix A:** Evaluation test

**Welke effecten heeft muziek op eetgedrag?**

***Deel 1***

**1 Inleiding**

In Nederland is er de afgelopen jaren een stijgende lijn in overgewicht te zien. Overgewicht wordt vaak onderverdeeld in matig en ernstig overgewicht. Volgens het Centraal Bureau voor de Statistiek (CBS) kampte in 2007 ruim 45% van de Nederlanders boven de 20 jaar met overgewicht. Bovendien ging toen 11% van de volwassenen gebukt onder ernstig overgewicht. Om dit overgewicht tegen te gaan is het van belang te onderzoeken welke variabelen bij kunnen dragen aan het verminderen van de voedselinname.

Wansink (2004) heeft een model samengesteld waarin hij aangeeft welke indirecte factoren in de omgeving invloed hebben op de voedselconsumptie. Indirecte invloeden zijn volgens hem licht, geluid, muziek en geur, omdat deze factoren het simpelweg comfortabeler en plezieriger maken voor een persoon om meer tijd te spenderen aan eten en men daardoor meer eet. Kotler deed een studie in verschillende restaurants (1973) met zowel gezond als minder gezond eten op hun menukaarten om te kunnen controleren voor het soort aanbod in menu's. Uit de studie bleek dat het meeste voedsel wordt geconsumeerd in restaurants waar de sfeer het prettigst is.

Uit eerder onderzoek blijkt dat muziek invloed heeft op de hoeveelheid voedsel die geconsumeerd wordt (Berlyne, 1971). Muziek is volgens Berlyne in te delen in muziek die hoge arousal<sup>1</sup> of lage arousal veroorzaakt. Hij definieert muziek die hoge arousal veroorzaakt als luide, onregelmatige en moeilijk te voorspellen muziek met een hoog tempo. Muziek die een lage arousal veroorzaakt definieert hij als zachte, monotone, erg voorspelbare muziek met een laag tempo. Het experiment in dit verslag sluit aan op dat van Berlyne en richt zich op het tempo van muziek. Er wordt gekeken welk effect dit heeft op de hoeveelheid voedsel die geconsumeerd wordt.

Uit latere studies blijkt dat het muziektempo een invloed heeft op de manier waarop mensen de sfeer ervaren (bijv. in een restaurant). Muziek met een hoog tempo verhoogt fysiologische arousal bij consumenten (Vanderark & Ely, 1993)

---

<sup>1</sup> Arousal: opwindend



waardoor de sfeer als minder plezierig wordt ervaren. Muziek met een laag tempo zorgt voor fysiologische ontspanning, waardoor de sfeer als plezierig wordt ervaren (Milliman, 1986). In het onderzoek in dit verslag wordt daarom aangenomen dat arousal een mediator variabele<sup>2</sup> is.

Caldwell en Hibbert (2002) deden onderzoek naar de relaties tussen muziektempo, muziekvoorkeur en eetgedrag. Hun onderzoek vond plaats in een fast food restaurant. Ze keken onder andere naar het verband tussen muziekvoorkeur en tijdbesteding in het restaurant, naar het effect van muziektempo op de hoeveelheid geld die werd uitgegeven en naar welk effect muziektempo had op de eetervaring en de intentie om terug te komen naar het restaurant. Uit het onderzoek bleek dat bij langzame muziek mensen langer in het restaurant bleven. Ook bleek de muziekvoorkeur samen te hangen met de tijd die men spendeerde aan het verblijf in het restaurant. Verder bleek dat bij langzamere muziek meer geld werd uitgegeven – voor het eten en het drinken – dan bij snelle muziek. Dit verschil bleek significant<sup>3</sup>. Dit effect kan volgens de onderzoekers echter ook zijn beïnvloed door factoren als het type restaurant, het bedieningssysteem of de specifieke dagen waarop het onderzoek is uitgevoerd. Als muziektempo echter invloed heeft op de uitgaven in een restaurant kan men aannemen dat het ook een invloed heeft op de hoeveelheid eten en drinken die men tot zich neemt.

### **Geef nu antwoord op Opgave 1.**

---

<sup>2</sup> Mediator variabelen: variabelen die de relatie tussen twee andere variabelen verklaren

<sup>3</sup> Significant: statistische relevantie: een verschil in de resultaten die met grote waarschijnlijkheid niet door toeval tot stand is gekomen.

## **2 Hypothesen**

Om de onderzoeksvraag te beantwoorden werden de volgende hypothesen (H) gesteld. Een hypothese is een verwachting ten aanzien van de resultaten.

H1: Het muziektempo beïnvloedt de hoeveelheid voedsel die geconsumeerd wordt.

H2: Bij langzame muziek wordt er meer geconsumeerd.

H3: Het muziektempo beïnvloedt de hoeveelheid gezond voedsel die geconsumeerd wordt.

H4: Het effect van muziektempo wordt mede beïnvloed door arousal en (daaraan gekoppeld) sfeer.

**Geef nu antwoord op Opgave 2.**

### **3 Methode**

Aan dit experiment hebben 54 studenten van de Universiteit Twente meegedaan. Daarvan waren er 15 man en 39 vrouw, de gemiddelde leeftijd was 21.97 jaar ( $SD^4 = 2.58$ ). De deelnemers werden over verschillende groepen (of: condities<sup>5</sup>) verdeeld.

De studenten werd verteld dat zij mee zouden doen aan een zogenaamde focusgroep ter verbetering van de campus van de Universiteit Twente. De deelnemers uit elke conditie werden hiervoor nog een keer opgesplitst in subgroepjes van zes mensen. De focusgroepbijeenkomsten vonden plaats in een ruimte op de campus van de Universiteit Twente, telkens om 15.45 uur. Deze tijd is gekozen omdat ze gezien kan worden als 'snack-tijd'. Er was een tafel in de ruimte aanwezig die groot genoeg was om met zeven personen (een confederate<sup>6</sup> en zes deelnemers) aan te zitten en voldoende stoelen. Tijdens de gesprekken was er voedsel binnen handbereik van de deelnemers op de tafel aanwezig. Verder was drinken aanwezig om te voorkomen dat de deelnemers door een gebrek aan drinken minder zouden eten. De muziek werd afgespeeld op een IBM laptop, waarbij het volume gedurende het experiment constant werd gehouden. Tevens waren er twee camera's aanwezig, zodat het eetgedrag van de deelnemers eventueel nog achteraf geanalyseerd kon worden.

De deelnemers werden bij binnenkomst verteld dat het voedsel er stond om hen te bedanken voor hun deelname. Het bestond uit gezond eten (stukken komkommer in verschillende groottes, druiven en cherrytomaatjes) en ongezond eten (verschillende snoepjes, gevulde koeken en biscuitjes). Hierdoor was er voor elk wat wils op tafel. Verder was er water en thee aanwezig. De deelnemers moesten zelf het eten pakken, omdat zij zelf de inname moesten bepalen. Per subgroepje heeft een assistent voor- en achteraf het aantal voedsleenheden geteld (aantal stukken komkommer, tomaatjes, snoepjes en koekjes), om zo te kunnen berekenen hoeveel er per conditie gegeten was. Deze tellingen werden aan de hand van de video opnames gecontroleerd.

---

<sup>4</sup> SD: standaarddeviatie.

<sup>5</sup> Condities: groepen proefpersonen die op een variabele verschillen; hierdoor kan men zien of de manipulatie van deze variabele tot verschillende resultaten leidt.

<sup>6</sup> Confederate: iemand die bij het onderzoek is betrokken. In dit geval degene die de discussie over de verbetering van de campus leidde.

Er werd uitgelegd dat de focusgroep maximaal 30 minuten zou duren, waarbij eerst algemene vragen over de campus werden gesteld, gevolgd door vragen over de sfeer en de leefbaarheid op de campus. Na het gesprek werd gevraagd aan de deelnemers om een vragenlijst in te vullen. In deze vragenlijst werd gevraagd naar demografische gegevens (bijv. leeftijd), de muziekvoorkeuren van de deelnemers en wat zij van het aangeboden eten vonden. Verder zijn vragenlijsten afgenomen waarmee de ervaren plezierigheid van de sfeer tijdens de bijeenkomst en de ervaren arousal werden gemeten. Ook heeft elke deelnemer in de vragenlijst zijn of haar geslacht aangegeven, zijn of haar lengte en gewicht – om de BMI-score te kunnen berekenen –, en of men op dieet is. De BMI-score, het dieet en wat men van de muziek en van het eten vond zijn als covariabelen<sup>7</sup> in de analyses meegenomen.

Voor de condities die aan muziek blootgesteld werden gold dat er gedurende de  $\pm 30$  minuten van het experiment muziek aanwezig moest zijn. De liedjes die gebruikt zijn, zijn door middel van Sony Sound Forge 9.0 aangepast qua beats per minuten (BPM), zodat er snelle (94 BPM) en langzame muziek (60 BPM) afgespeeld kon worden zonder dat de muziekgenres anders waren. De BPM zijn overgenomen van onderzoek van Milliman (1986). De muziek viel onder de categorie popmuziek, waarbij zang aanwezig was.

**Geef nu antwoord op Opgave 3.**

**Lever Deel 1 van de tekst in, als je klaar bent.**

---

<sup>7</sup> Covariabelen: covariabelen zijn variabelen die resultaten mogelijk kunnen beïnvloeden. Bijvoorbeeld, als mensen in dit onderzoek toevallig op dieet zijn eten ze minder. Door tijdens het onderzoek vast te stellen of deelnemers op dieet zijn of niet kan achteraf statistisch worden getoetst of dit de resultaten heeft beïnvloed.

## ***Deel 2***

### **4 Resultaten**

In Tabel 1 staat de gemiddelde hoeveelheid voedsel (aantal stukjes komkommer, snoepjes etc.) per conditie en per focusgroep. Per conditie zijn de standaarddeviaties aangegeven. Verder staan in Tabel 1 de gemiddelde scores op de arousal schaal en de plezier schaal per conditie en per focusgroep en de bijhordende standaarddeviaties aangegeven.

Op basis van een univariate variantieanalyse werd geconstateerd dat er een significant verschil<sup>8</sup> was in het eetgedrag tussen de condities. In de controle conditie werd meer gegeten dan in de snelle muziek conditie. Dit verschil was significant. In de langzame muziek conditie werd ongeveer evenveel gegeten als in de controle conditie en meer dan in de snelle muziek conditie. Dit laatste verschil was significant. Er was geen significant verschil tussen de controle conditie en de langzame muziek conditie.

Vanwege de aanname dat plezier en arousal als mediator variabelen zouden optreden is gekeken naar de effecten van deze variabelen. Als naar de plezier-schaal werd gekeken, kon geconstateerd worden dat er geen significant verschil tussen de condities was in de score op de plezier schaal. Hetzelfde goldt voor de score op de arousal schaal.

**Geef nu antwoord op Opgave 4 (met behulp van de onderstaande tabel en grafiek).**

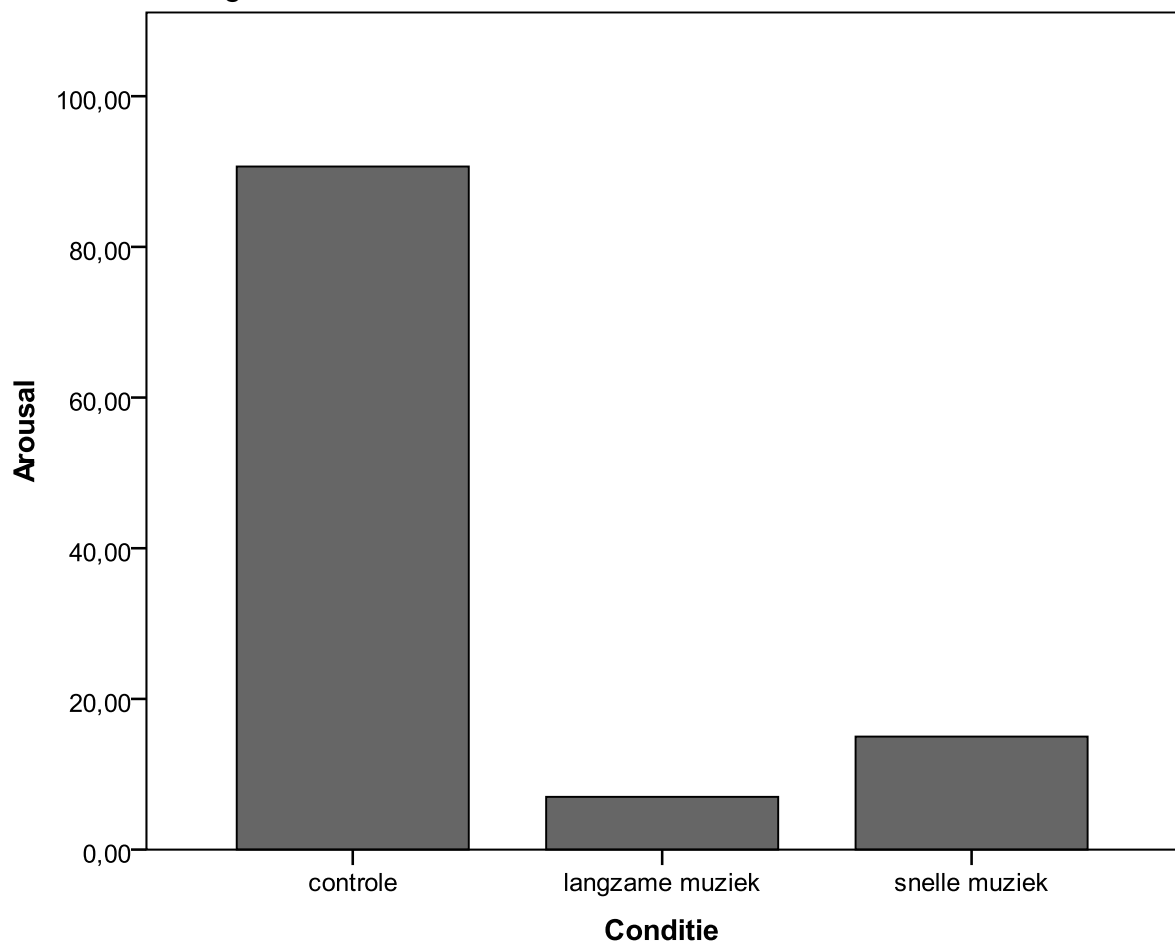
---

<sup>8</sup> Significant verschil: statistisch relevant; het verschil tussen de condities is dan niet toevallig

**Tabel 1**

| Conditie                  | Hoeveelheid eten | Plezier | Arousal |
|---------------------------|------------------|---------|---------|
| <u>controlegroep</u>      |                  |         |         |
| <i>focusgroep 1</i>       | 75               | 25      | 18      |
| <i>focusgroep 2</i>       | 76               | 27      | 12      |
| <i>focusgroep 3</i>       | 78               | 25      | 15      |
| <i>gemiddeld</i>          | 76               | 26      | 15      |
| <u>langzame muziek</u>    |                  |         |         |
| <i>focusgroep 4</i>       | 18               | 23      | 7       |
| <i>focusgroep 5</i>       | 11               | 28      | 9       |
| <i>focusgroep 6</i>       | 15               | 26      | 5       |
| <i>gemiddeld</i>          | 15               | 26      | 7       |
| <u>snelle muziek</u>      |                  |         |         |
| <i>focusgroep 7</i>       | 80               | 26      | 85      |
| <i>focusgroep 8</i>       | 84               | 26      | 96      |
| <i>focusgroep 9</i>       | 79               | 31      | 91      |
| <i>gemiddeld</i>          | 81               | 28      | 91      |
| <u>totaal</u>             |                  |         |         |
| <i>gemiddeld</i>          | 57               | 26      | 38      |
| <i>standaard deviatie</i> | 32               | 2       | 40      |

*Noot: De hoeveelheid eten wordt in stukken aangegeven, de plezier schaal liep van 1 tot 50 (1=weinig plezier, 50=veel plezier), de arousal schaal van 1- 100 (1=laag arousal, 100=hoog arousal)*



## **5 Conclusie en Discussie**

In dit onderzoek is ondersteuning gevonden voor de hypothese dat er bij langzame muziek meer gegeten wordt dan bij snelle muziek. Verder blijkt uit de resultaten dat in de snelle muziek conditie minder gegeten werd dan in de controle conditie waar helemaal geen muziek gespeeld werd. Wanneer op groepsniveau de resultaten uit de experimentele condities met elkaar vergeleken werden, lieten de resultaten verder zien dat geen duidelijk verschil tussen de langzame muziek conditie en de controle conditie was.

In deze studie is gekeken naar de effecten van de omgevingsfactor muziek op het eetgedrag. Het muziekexperiment laat zien dat er waarschijnlijk meer gegeten wordt in restaurants waar langzame muziek gespeeld wordt dan in restaurants waar snelle muziek gespeeld wordt. Daarnaast is duidelijk dat mensen ook meer eten als er in een restaurant helemaal geen muziek aanwezig is in vergelijking met als er snelle muziek te horen is.

Uiteraard zijn er ook een aantal beperkingen aan dit experiment. De deelnemers durfden misschien niet om het eten te pakken. Men kon niet alles pakken zonder het gesprek te 'storen'. Verder is het mogelijk dat de invloed van muziek op eetgedrag anders is dan gedacht. Ondanks dat de langzame muziek in een pilot-study<sup>9</sup> langzamer werd bevonden en de snelle muziek sneller, zou het kunnen zijn dat de waarden van 60 BPM voor langzame muziek en 94 BPM voor snelle muziek verouderd zijn. Als gekeken wordt naar het tempo in popmuziek blijkt dat veel liedjes tegenwoordig meer een waarde richting de 110 BPM hebben. Als blijkt dat de waarden verouderd zijn, zouden liedjes met hogere BPM meer opwinding kunnen opwekken, waardoor wellicht andere resultaten met betrekking tot het eetgedrag zouden zijn gevonden. Onderzoek naar het effect van verschillende tempo's op arousal zou hier uitsluitsel over kunnen geven. Ook andere factoren kunnen invloed hebben gehad. Aan dit onderzoek hebben alleen studenten (rond 22 jaar) deelgenomen. Verder vonden alle focusgroepen in dezelfde kamer plaats. Daarom zou in een volgend experiment bijvoorbeeld gekeken kunnen worden naar de invloed van omgeving (bijv. kleur van de muren) en leeftijd op voedselconsumptie.

---

<sup>9</sup> Pilot-study: onderzoek dat vooraf wordt uitgevoerd om bijvoorbeeld de gebruikte instrumenten (vragenlijsten, manipulaties, ...) uit te proberen.

Uit het experiment blijkt dus dat muziektempo invloed heeft op de voedselconsumptie. Een plek waar de resultaten goed kunnen worden toegepast is in een “Eet-zoveel-je-wilt” restaurant waar men voor één prijs zoveel kan eten als men wilt. Vanuit het perspectief van de restauranteigenaar wil je er eigenlijk voor zorgen dat men juist zo min mogelijk eet. Aangezien de resultaten laten zien dat bij de snelle muziek er minder gegeten wordt, kan deze kennis goed worden toegepast in dergelijke restaurants. Gezien de stijgende problemen met overgewicht (CBS 2008), kunnen deze resultaten echter ook van maatschappelijk belang zijn.

**Geef nu antwoord op Opgave 5.**



## **Opdracht**

Bij deze opdracht vind je een wetenschappelijke tekst over de invloed van muziek op eetgedrag. Het artikel is gebaseerd op een afstudeeropdracht van een student van de Universiteit Twente. Over de tekst ga je een aantal vragen beantwoorden. Deze vragen staan op aparte antwoordbladen. Je hebt hier 90 minuten de tijd voor.

Voor deze opdracht is de oorspronkelijke tekst ingekort en op sommige punten met opzet aangepast. Om die reden kan het zo zijn dat sommige in de tekst vermelde gegevens en resultaten niet kloppen. De samenvatting en de literatuurlijst zijn verwijderd, omdat deze voor de opdracht niet van belang zijn. Sommige woorden, die je wellicht niet kent, worden in voetnoten uitgelegd.

Bestudeer de tekst en beantwoord de vragen die je tegenkomt in de volgorde zoals ze worden gepresenteerd (1,2,3 ...). Gebruik hiervoor de antwoordbladen. Als je met een antwoordblad klaar bent leg je hem omgedraaid op de hoek van je tafel. Begin daarna pas met de volgende opgave. De proefleider zal langskomen om het antwoordblad op te halen. Voor het onderzoek is het van belang dat je het op deze manier doet en niet terugkijkt naar vorige opgaven als je aan een nieuwe opgave bent begonnen. Probeer steeds de vragen volledig te beantwoorden.

De tekst is gesplitst in Deel 1 en Deel 2. Steek je hand op als je met Deel 1 en de bijbehorende opgaven klaar bent. Je ontvangt dan Deel 2.

Bij sommige opgaven word je gevraagd om informatie in de tekst op te zoeken, in andere opgaven moet je aangeven of informatie in de tekst correct is of niet. Gebruik eventueel de regelnummers in de kantlijn van de tekst om naar tekstgedeelten te verwijzen. Zorg dat je alle vragen beantwoord.

Als je klaar bent lever je de tekst en antwoordbladen in bij de proefleider.

Vul nu eerst dezelfde code die je voor de online-vragenlijst hebt gebruikt aan de bovenkant van het blad in. Combineer hiervoor de laatste letter van je voornaam, de laatste letter van je achternaam en de dag en de maand van je geboortedatum. Zo wordt bijv. Katja Hoffman, geboren op 23 augustus: AN2308. Vul de code ook op alle antwoordbladen in!

**Begin niet met de opgaven voordat je de instructie krijgt om dit te doen**

## Deel 1

### **Opgave 1: Inleiding – Identificeer de variabelen**

1.1 Benoem hieronder alle variabelen uit de verschillende onderzoeken die worden besproken in de inleiding.

1.2 Geef hieronder drie voorbeelden van variabelen uit de inleiding die door onderzoekers zelf werden gemanipuleerd.

1.3 Geef ook drie voorbeelden van variabelen uit de inleiding die onderzoekers gebruikten om hun conclusies op te baseren.

**Ga nu terug naar de tekst naar onderdeel 2 Hypothesen.**

**Opgave 2: Hypothesen - Kijk naar de hypothesen. Wat denk je?**

2.1 Welke hypothesen zijn relevant voor de onderzoeksvraag die in de inleiding aangeduid wordt? Leg bij elke hypothese uit waarom hij wel of niet relevant is voor de onderzoeksvraag. (Let op: De titel is niet de onderzoeksvraag)

Relevant voor de onderzoeksvraag:

Niet relevant voor de onderzoeksvraag:

2.2 Omschrijf welke hypothese ontbreekt in het rijtje. Motiveer je antwoord.

**Ga nu terug naar de tekst naar onderdeel 3 Methode**

**Opgave 3: Methode**

3.1 In de tekst staat niet duidelijk vermeld welke groepen (of: condities) met elkaar vergeleken zijn. Welke zullen dat geweest zijn volgens jou? Beschrijf ze kort.

3.2 Welke variabele zou door de onderzoekers zelf gemanipuleerd moeten worden om hypothese 3 (H3) te kunnen toetsen? En welke condities zou je met elkaar vergelijken om H3 te toetsen?

3.3 Stel dat je hypothese H3 onderzoekt. Op welke variabele(n) zou je je conclusies dan het liefst baseren?

3.4 Hoe wordt gemeten of de hoeveelheid voedsel tussen de condities verschilt?

3.5 Er is een probleem met de manier waarop de hoeveelheid voedsel die geconsumeerd is wordt gemeten. Weet jij welk probleem?

3.6 Kun je voor dit onderzoek een betere manier bedenken om te meten of de hoeveelheid voedsel die geconsumeerd wordt tussen de condities verschilt?

**Lever nu Deel 1 van de tekst in door je hand op te steken en begin met Deel 2. Lees voor de volgende opgaven eerst onderdeel 4 Resultaten in de tekst goed door.**

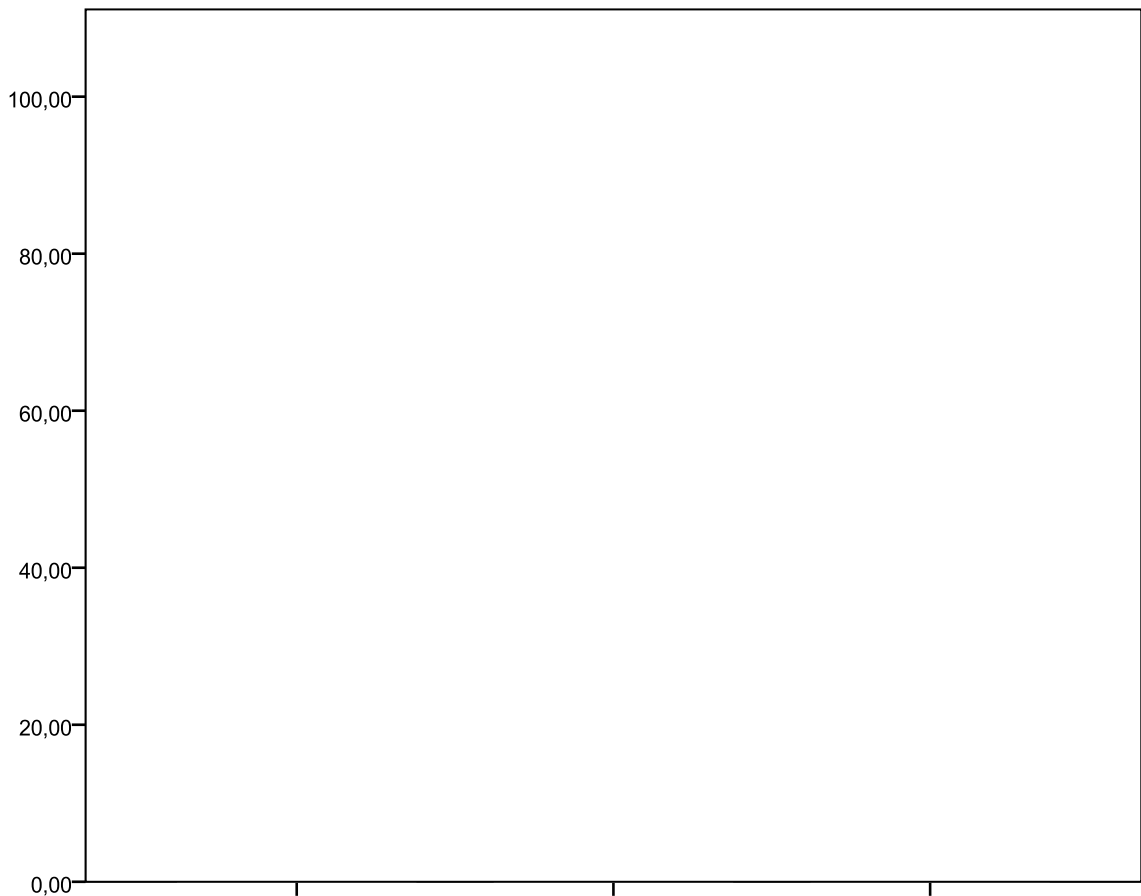
**Deel 2**

**Opgave 4: Resultaten**

4.1 Kijk naar de resultaten in Tabel 1. Worden de resultaten uit de tekst op een correcte wijze weergegeven? Motiveer je antwoord.

4.2 Vergelijk de tabel en de grafiek. Valt je iets op?

4.3 Stel dat Hypothese 3 (H3: Het muziektempo beïnvloed de hoeveelheid gezond voedsel die geconsumeerd wordt) ook onderzocht werd en de resultaten waren als volgt: In de snelle muziek conditie wordt significant meer gezond voedsel geconsumeerd dan in de controle conditie en de langzame muziek conditie. Er werd geen significant verschil tussen de controle conditie en de langzame muziek conditie gevonden. Schets hieronder hoe een grafiek uit zou kunnen zien. Gebruik de grafiek uit de tekst als voorbeeld.



**Ga nu terug naar de tekst naar onderdeel 5 Conclusie en discussie.**



**5: Conclusie en Discussie**

(Ga voor opgave 5 uit van de resultaten zoals die in de TEKST staan )

5.1 Vergelijk de conclusies en de resultaten. Benoem de conclusies die worden getrokken. Zijn deze correct? Motiveer je antwoord.

5.2 Welke conclusie ontbreekt er gezien de resultaten? Motiveer je antwoord.

5.3 Welk advies zou je een gewoon restaurant (dus geen “Eet-zoveel-als-je-wilt” restaurant) geven op basis van de resultaten van dit onderzoek? Motiveer je antwoord.

## Bedankt voor je medewerking.

**Appendix B:** Concept version of the scientific text understanding assignment

| Subscale           | Part of scientific paper | Question related to IST  |
|--------------------|--------------------------|--|
| Identify variables | Introduction             | <p>Identify the variables in the text.</p> <p>1: Which variables should be included in the study to answer the research question?</p> <p>Which variables are not necessary to answer the research questions?</p> <p>What do you think of the variables?</p> <p>Could you think of additional variables?</p> <p>Which variables should be excluded?</p> <p>3: Which variable may have an influence on variable X?</p> <p>Which variable may be influence by variable X?</p> <p>How do the variables influence each other?</p> <p>13: Which variable is the control variable? Is there a control variable (which?). If not, think of one. If yes, is there a better one?</p> <p>14: Which variable is the dependent variable? Is there a dependent variable (which?). If not, think of one. If yes, is there a better one?</p> <p>15: Which variable is the independent variable? Is there an independent variable (which?). If not, think of one. If yes, is there a better one?</p> <p>18, 19, 20, 30, 31 &amp; 32: cf. 13, 14 &amp; 15</p> <p>35: cf. 1 &amp; 3</p> |
| Make hypotheses    | Introduction/hypotheses  | <p>4: Which hypotheses help to answer the research question?</p> <p>Which hypotheses do not help to answer the research question?</p> <p>Do you think additional hypotheses should be tested? If Yes, which and why? If not, why)</p> <p>6, 8, 16, 27, 48: cf. 4</p> <p>8: Which hypotheses are tested in the study?</p> <p>Which hypotheses can you find in the text?</p>   |

|                    |                         |  |
|--------------------|-------------------------|--|
|                    |                         | <p>What do you think about the hypotheses?</p> <p>Which hypotheses do you think miss?</p> <p>17, 29: cf. 12</p> <p>34: Are the hypotheses based on the chosen variables? Which are? Which are not? Which hypotheses miss with regard to the variables?</p> <p>Look at the hypotheses. Which are good? Which are not good?</p> <p>Would you change something?</p>   |
| Define operational | Method                  | <p>2: How is hypothesis X tested in this study? What do you think about the method? Would you change something? Would you choose another method? If yes which?</p> <p>How are the hypotheses in this study tested? Would you change something? Would you choose another method? If yes, which?</p> <p>Think about other methods that could be used.</p> <p>7, 37, 39, 44: cf. 2</p> <p>22: Is there something wrong about the methods used?</p> <p>Do the methods help to answer the research questions?</p> <p>Do the methods test the hypotheses?</p> <p>23: How can you see if hypothesis X is correct?</p> <p>What would answer the research question?</p> <p>26: cf. 23</p> <p>33: What is tested with method X?</p> <p>Are there unnecessary tests included in the study?</p> <p>41: What do you think about the method in general? About the respondents? About the time planning? About the setting? Would you improve/change/add something?</p> |
| Design testing     | Method/test description | <p>10: Are the methods used to test the hypothesis correct?</p>  |

|                                    |                   |   |
|------------------------------------|-------------------|---|
|                                    |                   | <p>Have you any other ideas how the hypothesis could be tested?</p> <p>For hypothesis X no test is described. Can you think of a method to test hypothesis X?</p> <p>21, 24, 43, 46, 50, 52: cf. 10</p> <p>38: Which of the methods leads to the reliable results?</p> <p>Which of the methods is useful for answering the research question? Which is not? Can you improve them?</p> <p>42: How can the relation between variable X and variable Y be tested?</p> <p>49: Do you have ideas how hypothesis X could be tested?</p>   |
| <p>Interpret graphics and data</p> | <p>Results</p>    | <p>5: Are all graphics correct?</p> <p>Which of the graphics are incorrect?</p> <p>Look at the table and the graph in the text. Do you think they match? Why/why not?</p> <p>Can you explain how the graphic should look like based on the results from the table?</p> <p>9: Look at the graphic. What does it show? Is the graphic correct?</p> <p>Is the graphic useful to understand the results?</p> <p>Is the graphic necessary to understand the results?</p> <p>Is the graphic necessary to answer the research question?</p> <p>What other graphic/table should be included in the results?</p> <p>11, 45, 51: cf. 9</p> <p>25, 36: cf. 5</p> <p>28: Are the conclusions about the graphic/table correct? Are there additional conclusions that can be made? If not, why? If yes, why) (formulate them)</p> <p>40, 47: cf. 28</p> |
| <p>Conclusion</p>                  | <p>Conclusion</p> | <p>W/O: Which conclusions are made in the text?</p>   |

|                |            |  |
|----------------|------------|--|
|                |            | <p>Are the conclusions correct? If yes, why? If not, why not? Can you correct them?</p> <p>Are all possible conclusions included in the text?</p> <p>Are the conclusions made necessary to answer the research question?</p> <p>Do the conclusions answer the hypothesis?</p> <p>Are all hypothesis answered? Are all important results mentioned? If not, answer them based on the results.</p> <p>IO: Are enough information given in the text to make the presented conclusions?</p> <p>Are the conclusions based on the results? If not, why?</p> <p>Would you drop conclusions?</p> <p>WW/WO: Are the conclusions fully based on the results?</p> <p>Are some of the conclusions based on assumptions?</p> <p>Are all hypotheses confirmed or refuted based on the results?</p> |
| Interpretation | Conclusion | <p>Are the results interpreted correctly? Why/why not?</p> <p>Would you come to a different interpretation?</p> <p>Are all possible interpretations included in the text?</p> <p>Does the interpretation answer the research question? If not, why? Can you write a better/correct interpretation? Can you write an interpretation that gives an answer to the research question?</p>  |

**Appendix C:** Evaluation scheme for the scientific text understanding assignment

| Opgave | 0 punten   | 1 punt             | 2 punten           | 3 punten          |
|--------|--|--------------------|--------------------|-------------------|
| 1.1    | <p>Benoem hieronder alle variabelen uit de verschillende onderzoeken die worden besproken in de inleiding.</p> <p>De in de inleiding genoemde variabelen zijn:</p> <ul style="list-style-type: none"> <li>- overgewicht (ernstig en/of matig)</li> <li>- voedselconsumptie/voedselinname/hoeveelheid (geconsumeerd) voedsel/hoeveelheid (geconsumeerd) eten (en drankjes)</li> <li>- licht</li> <li>- geluid</li> <li>- geur</li> <li>- muziek (luid en/of zacht, onregelmatig en/of monotoon, moeilijk te voorspellen en/of voorspelbaar)</li> <li>- voedsel/eten (gezond en/of ongezond/minder gezond)</li> <li>- soort aanbod</li> <li>- sfeer/ervaren sfeer (gezellig en/of ongezellig, plezierig en/of onplezierig/minder plezierig)</li> <li>- arousal/fysiologisch arousal (hoog en/of laag)</li> <li>- ontspanning/fysiologische ontspanning</li> <li>- muziektempo (hoog en/of laag)</li> <li>- muziekvoorkeur</li> <li>- eetgedrag</li> <li>- tijd in het restaurant/tijdbesteding</li> <li>- gelduitgave/hoeveelheid uitgegeven geld</li> <li>- eetervaring</li> <li>- intentie om terug te komen</li> <li>- type restaurant</li> <li>- bedieningssysteem</li> <li>- (specifieke) dagen waarop het onderzoek is uitgevoerd</li> </ul> <p><i>De deelnemers krijgen voor elk hoofdpunt een punt. Bv. leveren de variabelen "muziektempo" en "muziekvoorkeur" twee punten op. Voor de antwoorden "ernstig overgewicht, matig overgewicht" krijgt de deelnemer echter maar een punt. Herhalingen worden maar een keer beoordeeld. Hetzelfde geldt voor het gebruik van synonieme. Bv. de antwoorden "voedselconsumptie" en "voedselinname" leveren een punt op.</i></p> <p><i>Als de deelnemer een synoniem van een woord uit de bovenstaande tabel gebruikt en dit synoniem zelf niet in de bovenstaande tabel te vinden is krijgt de deelnemer een punt. Bv. voor het antwoord "soort restaurant" in plaats van "type restaurant" krijgt de deelnemer een punt. Als de deelnemer zowel "soort restaurant" als ook "type restaurant" gebruikt, levert dit echter maar een punt op.</i></p> <p><i>De antwoord "snelle en langzame muziek" is hetzelfde als "muziektempo". Als allebei genoemd worden geeft dit dus maar een punt.</i></p> |                    |                    |                   |
|        | <b>Geen antwoord<br/>of<br/>0-2 correct</b>  | <b>3-5 correct</b> | <b>6-8 correct</b> | <b>9+ correct</b> |
| 1.2    | <p>Geef hieronder drie voorbeelden van variabelen uit de inleiding die door onderzoekers zelf werden gemanipuleerd.</p> <p>Variabelen uit de inleiding die door onderzoekers zelf werden gemanipuleerd (onafhankelijke variabele):</p> <ul style="list-style-type: none"> <li>- licht</li> <li>- geluid</li> </ul>   |                    |                    |                   |

|  |   |  |                  |                  |                  |
|--|---|--|------------------|------------------|------------------|
|  | <ul style="list-style-type: none"> <li>- geur</li> <li>- muziek (luid en/of zacht, onregelmatig en/of monotoon, moeilijk te voorspellen en/of voorspelbaar)</li> <li>- voedsel/eten (gezond en/of ongezond/minder gezond)</li> <li>- soort aanbod</li> <li>- muziektempo (hoog en/of laag)</li> <li>- muziekvoorkeur</li> <li>- type restaurant</li> <li>- bedieningssysteem</li> <li>- (specifieke) dagen waarop het onderzoek is uitgevoerd</li> </ul> <p><i>De deelnemers krijgen voor elk hoofdpunt een punt. Bv. leveren de variabelen “muziektempo” en “muziekvoorkeur” twee punten op. Voor de antwoorden “ernstig overgewicht, matig overgewicht” krijgt de deelnemer echter maar een punt. Herhalingen worden maar een keer beoordeeld. Hetzelfde geldt voor het gebruik van synonieme. Bv. de antwoorden “voedselconsumptie” en “voedselinname” leveren een punt op.</i></p> <p><i>Als de deelnemer een synoniem van een woord uit de bovenstaande tabel gebruikt en dit synoniem zelf niet in de bovenstaande tabel te vinden is krijgt de deelnemer een punt. Bv. voor het antwoord “soort restaurant” in plaats van “type restaurant” krijgt de deelnemer een punt. Als de deelnemer zowel “soort restaurant” als ook “type restaurant” gebruikt, levert dit echter maar een punt op.</i></p> <p><i>De antwoord “snelle en langzame muziek” is hetzelfde als “muziektempo”. Als allebei genoemd worden geeft dit dus maar een punt.</i></p> <table border="1" data-bbox="347 1034 1418 1137"> <tr> <td data-bbox="347 1034 619 1137"><b>Geen antwoord</b><br/>of<br/><b>0 correct</b></td> <td data-bbox="619 1034 927 1137"><b>1 correct</b></td> <td data-bbox="927 1034 1177 1137"><b>2 correct</b></td> <td data-bbox="1177 1034 1418 1137"><b>3 correct</b></td> </tr> </table> | <b>Geen antwoord</b><br>of<br><b>0 correct</b> | <b>1 correct</b> | <b>2 correct</b> | <b>3 correct</b> |
| <b>Geen antwoord</b><br>of<br><b>0 correct</b> | <b>1 correct</b>  | <b>2 correct</b>                               | <b>3 correct</b> |                  |                  |
| 1.3  | <p>Geef ook drie voorbeelden van variabelen uit de inleiding die onderzoekers gebruikten om hun conclusies op te baseren.</p> <p>Variabelen uit de inleiding die onderzoekers gebruikten om hun conclusies op te baseren (afhankelijke variabelen):</p> <ul style="list-style-type: none"> <li>- Overgewicht (ernstig en/of matig)</li> <li>- voedselconsumptie/voedselinname/hoeveelheid (geconsumeerd) voedsel/hoeveelheid (geconsumeerd) eten (en drankjes)</li> <li>- sfeer/ervaren sfeer (gezellig en/of ongezellig, plezierig en/of onplezierig/minder plezierig)</li> <li>- arousal/fysiologisch arousal (hoog en/of laag)</li> <li>- ontspanning/fysiologische ontspanning</li> <li>- eetgedrag</li> <li>- tijd in restaurant/tijdbesteding</li> <li>- gelduitgave/hoeveelheid uitgegeven geld</li> <li>- eetervaring</li> <li>- intentie om terug te komen</li> </ul> <p><i>De deelnemers krijgen voor elk hoofdpunt een punt. Bv. leveren de variabelen “muziektempo” en “muziekvoorkeur” twee punten op. Voor de antwoorden “ernstig overgewicht, matig overgewicht” krijgt de deelnemer echter maar een punt. Herhalingen worden maar een keer beoordeeld. Hetzelfde geldt voor het gebruik van synonieme. Bv. de antwoorden “voedselconsumptie” en “voedselinname” leveren een punt op.</i></p> <p><i>Als de deelnemer een synoniem van een woord uit de bovenstaande tabel gebruikt en dit synoniem zelf niet in de bovenstaande tabel te vinden is krijgt de deelnemer een punt. Bv. voor het antwoord “soort restaurant” in plaats van</i></p>  |  |                  |                  |                  |

|            |   |                         |                         |                         |
|------------|---|-------------------------|-------------------------|-------------------------|
|            | <p><i>“type restaurant” krijgt de deelnemer een punt. Als de deelnemer zowel “soort restaurant” als ook “type restaurant” gebruikt, levert dit echter maar een punt op.</i></p> <p><i>In het onderzoek wordt ervan uitgegaan dat arousal (en hieraan gerelateerd sfeer) een mediatorvariabele is. Evenveel is arousal in een van de in de inleiding genoemde studies een variabele waarop onderzoekers hun conclusies baseren (een afhankelijke variabele).</i></p>   |                         |                         |                         |
|            | <p><b>Geen antwoord</b><br/>of<br/><b>0 correct</b></p>   | <p><b>1 correct</b></p> | <p><b>2 correct</b></p> | <p><b>3 correct</b></p> |
| <p>2.1</p> | <p>Welke hypothesen zijn relevant voor de onderzoeksvraag die in de inleiding aangeduid wordt? Leg bij elke hypothese uit waarom hij wel of niet relevant is voor de onderzoeksvraag. (Let op: De titel is niet de onderzoeksvraag.)</p> <p>De onderzoeksvraag die in de inleiding aangeduid wordt is: Hoe is muziektempo gerelateerd aan de hoeveelheid voedsel die geconsumeerd wordt? (r. 25-27)</p> <p><u>Relevante variabelen:</u><br/> H1: sluit direct aan bij de onderzoeksvraag (door duidelijk te maken dat in het onderzoek van een causaal effect wordt uitgegaan → muziektempo beïnvloed de hoeveelheid geconsumeerd voedsel</p> <p>H2: test een deel van de onderzoeksvraag, t.w. wat er bij een bepaald muziektempo (in dit geval langzame muziek) ten opzichte van de hoeveelheid eten die geconsumeerd wordt gebeurt (in dit geval “men eet meer”)</p> <p>H4: is relevant, omdat het de mediatorvariabele is waarvan met betrekking tot de onderzoeksvraag in de inleiding sprake is: R32-33: “In het onderzoek in dit verslag wordt daarom aangenomen, dat arousal een mediator variabele is.” Heeft indirect met de onderzoeksvraag te maken, omdat verondersteld wordt dat men door de verschillende niveaus van arousal die door snelle en langzame muziek worden opgeroepen een verschil in de hoeveelheid voedsel die geconsumeerd wordt tot stand komt.</p> <p>De redenering volgende redeneringen zijn voor alle hypothesen correct:</p> <ul style="list-style-type: none"> <li>- het testen van deze hypothese geeft antwoord op de onderzoeksvraag</li> <li>- deze hypothese specificeert de onderzoeksvraag</li> <li>- deze hypothese wordt getest om te zien hoe mensen op verschillende muziektempo’s reageren</li> <li>- door deze hypothese wordt het verschil in eetgedrag onderzocht</li> </ul> <p><i>De bovenstaande hypothesen H1, H2 en H4 zijn relevant voor de onderzoeksvraag. De redeneringen hiervoor kunnen verschillen. In de praktijk moet dus in elk geval apart worden beoordeeld of een redenering juist is of niet. Als oriëntatie kunnen de bovengenoemde redeneringen gebruikt worden, deze zijn of specifiek voor een hypothese of algemeen geldig.</i></p> <p><u>Niet relevante variabelen:</u><br/> H3: sluit niet aan bij de in de inleiding aangeduide onderzoeksvraag/ beantwoord niet de vraag of muziektempo aan de hoeveelheid voedsel dat geconsumeerd wordt gerelateerd is, hierbij is dus niet belangrijk onder welke omstandigheden men meer of minder eet, maar onder welke omstandigheden men voor gezond of ongezond eten kiest.</p> <p><i>De bovenstaande hypothese H3 is niet relevant voor de onderzoeksvraag. De</i></p> |                         |                         |                         |



|     |  |  |   |  |
|-----|--|--|---|--|
|     | <p><i>redenering hiervoor kan in formulering verschillen. In de praktijk moet dus in elk geval apart worden beoordeeld of een redenering juist is of niet. Als oriëntatie kan de bovengenoemde redenering gebruikt worden. Het moet een ieder geval duidelijk zijn dat H3 op basis van de in de inleiding gegeven informatie geen waarde aan de beantwoording van de onderzoeksvraag toevoegt.</i></p>   |  |   |  |
|     | <p><b>Geen antwoord</b><br/><i>of</i><br/><b>Geen correcte antwoord</b><br/><i>of</i><br/><b>Maar een correct antwoord zonder (correcte) redenering</b></p>  | <p><b>Een correcte antwoord en een correcte redenering</b><br/><i>of</i><br/><b>2 correcte antwoorden zonder (correcte) redenering</b></p>                             | <p><b>Twee correcte antwoorden en ten minste een correcte redenering</b><br/><i>of</i><br/><b>3 correcte antwoorden zonder (correcte) redenering</b></p>  | <p><b>3 correcte antwoorden en ten minste een correcte redenering</b><br/><i>of</i><br/><b>4 correcte antwoorden met of zonder (correcte) redenering</b></p> |
| 2.2 | <p>Omschrijf welke hypothese ontbreekt in het rijtje. Motiveer je antwoord.</p> <p>H2b mist (<i>hypothese moet niet H2b genoemd worden</i>): Bij snelle muziek wordt minder geconsumeerd.<br/>→Mist bij de hypothesen, omdat er alleen een uitspraak gedaan wordt over wat er bij het draaien van langzame muziek gebeurt. Zoals uit de inleiding naar voren komt, wordt in het onderzoek de voedselconsumptie met betrekking tot langzame <u>en</u> snelle muziek getoetst om de onderzoeksvraag te beantwoord.</p> <p><i>Er zijn een aantal hypothesen genoemd die theoretisch correct geformuleerd zijn, maar niets aan de beantwoording van de onderzoeksvraag toevoegen. Bv. zal men uit H2(Bij langzame muziek wordt meer geconsumeerd) kunnen redeneren, dat mensen die meer consumeren langer in een restaurant blijven en andersom. Dat leidt tot de hypothese: Bij langzame muziek blijft men langer in een restaurant. Voor dergelijke antwoorden krijgen deelnemers een punt, omdat de hypothese wel met betrekking tot de onderzoeksvraag (in het zelfde onderzoeksdesign) getoetst kan worden, maar niet echt aan de beantwoording van de onderzoeksvraag bijdraagt.</i></p> <p><i>Deelnemers die een hypothese en/of aangeven die correct is en iets aan de onderzoeksvraag toevoegt worden volgens de voorwaarden voor twee punten en voor drie punten beoordeeld.</i></p> |  |   |  |
|     | <p><b>Geen antwoord</b><br/><i>of</i><br/><b>fout antwoord</b><br/><i>of</i><br/><b>hypothese geformuleerd die een herhaling (bv. in andere woorden) van een andere hypothese is</b></p>   | <p><b>Hypothese geformuleerd die met betrekking tot de onderzoeksvraag theoretisch correct is, maar niets aan de beantwoording van de onderzoeksvraag toevoegt</b></p> | <p><b>Correcte hypothese geformuleerd, maar geen (correcte) redenering geformuleerd</b><br/><i>of</i><br/><b>alleen correcte redenering geformuleerd, zonder de hypothese (correct) te formuleren</b></p> | <p><b>Correcte hypothese en correcte redenering geformuleerd</b></p>   |

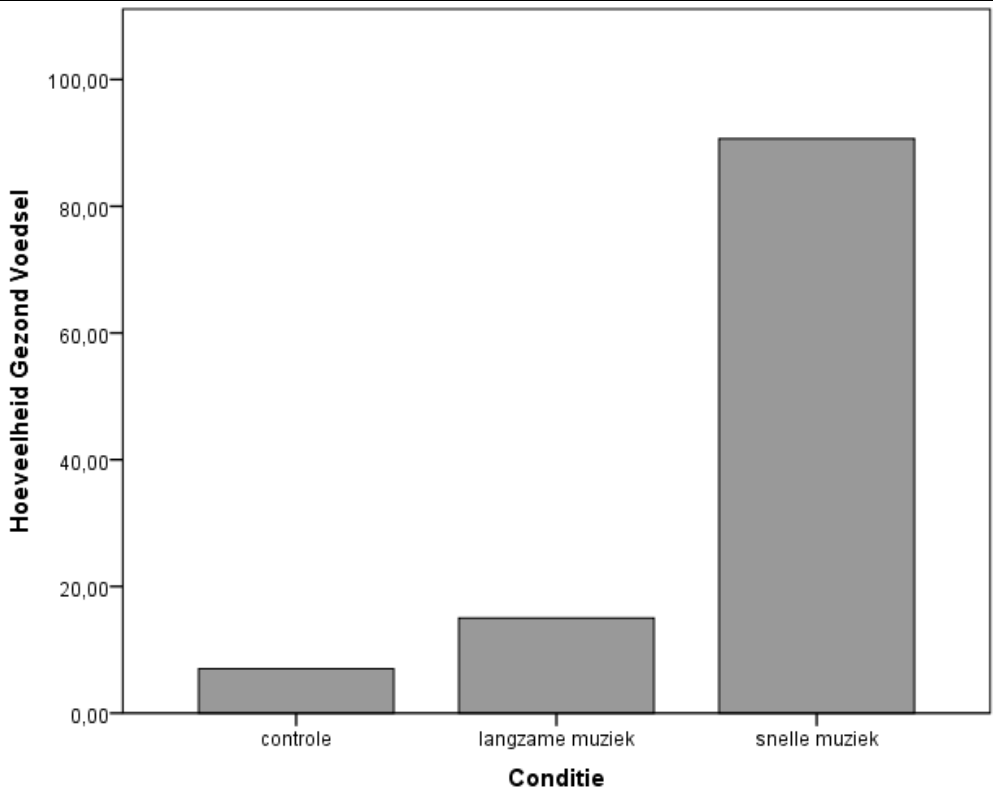
|     |  |   |  |  |
|-----|--|---|--|--|
| 3.1 | In de tekst staat niet duidelijk vermeld welke groepen (of: condities) met elkaar vergelken zijn. Welke zullen dat geweest zijn volgens jou? Beschrijf ze kort.  |   |  |  |
|     | Er zijn drie condities die getoetst worden:<br>- conditie 1=snelle muziek<br>- conditie 2=langzame muziek<br>- conditie 3=controlegroep/geen muziek  |   |  |  |
|     | <b>Geen antwoord<br/>of<br/>fout antwoord</b>  | <b>Een conditie<br/>genoemd<br/>Of<br/>Geeft aan dat BPM<br/>een rol speelt</b>   | <b>Twee condities<br/>genoemd</b>  | <b>Drie condities<br/>genoemd</b>  |
| 3.2 | Welke variabele zou door de onderzoekers zelf gemanipuleerd moeten worden om hypothese 3 (H3) te kunnen toetsen? En welke condities zou je met elkaar vergelijken om H3 te toetsen?  |   |  |  |
|     | Om H3 te testen zal natuurlijk ook het muziekt tempo gemanipuleerd worden, maar verder moet er nog het eten gemanipuleerd worden. Men moet erop letten dat het verschil tussen voedselconsumptie van gezond en van ongezond eten gemanipuleerd wordt. Hiervoor kan men verschillende condities met elkaar vergelijken. Of de onderzoekers vergelijken drie verschillende condities (langzame muziek, snelle muziek, controle) of zes condities (langzame muziek en gezond eten, langzame muziek en ongezond eten, snelle muziek en gezond eten, snelle muziek en ongezond eten, controlegroep en gezond eten en controlegroep en ongezond eten). |   |  |  |
|     | <i>De antwoorden op deze vragen lijken wellicht op elkaar. Hierbij is het belangrijk in te schatten of de deelnemer het onderzoek zo goed begrijpt dat een, twee of drie punten gepast zijn.</i>   |   |  |  |
|     | <b>Geen antwoord<br/>of<br/>fout antwoord</b>  | <b>Alleen het woord<br/>“muziekt tempo” of<br/>het woord<br/>“voedsel” of<br/>“gezond/ongezond”<br/>of “snel, langzaam,<br/>controle” genoemd,<br/>zonder precies aan<br/>te geven hoe deze<br/>variabele<br/>gemanipuleerd<br/>moeten worden en<br/>zonder verder op<br/>de condities in te<br/>gaan</b> | <b>De deelnemer<br/>noemt of het<br/>muziekt tempo<br/>of het voedsel<br/>dat<br/>gemanipuleerd<br/>moet worden<br/>en geeft aan<br/>hoe deze<br/>variabele<br/>gemanipuleerd<br/>wordt, er wordt<br/>niet verder op<br/>de condities<br/>ingegaan<br/>of<br/>Zowel<br/>“muziekt tempo”<br/>als ook<br/>“voedsel”<br/>genoemd,<br/>zonder aan te<br/>geven hoe deze<br/>variabelen<br/>precies<br/>gemanipuleerd</b> | <b>De deelnemer<br/>maakt<br/>duidelijk dat<br/>zowel het<br/>muziekt tempo,<br/>als ook het<br/>voedsel<br/>gemanipuleerd<br/>moeten<br/>worden, een<br/>geeft aan hoe<br/>deze<br/>gemanipuleerd<br/>moeten<br/>worden,<br/>hieruit wordt<br/>ook duidelijk<br/>welke<br/>condities er<br/>vergeleken<br/>worden</b> |

|     |   |   |   |  |
|-----|---|---|---|--|
|     |   |   | <p>worden, er wordt niet verder op de condities ingegaan</p> <p><i>of</i></p> <p>Het zijn condities aangegeven, die of duidelijk maken dat gezond en ongezond voedsel worden vergeleken, of dat dat drie verschillende muziektempo's worden vergeleken</p>        |  |
| 3.3 | <p>Stel dat je hypothese H3 onderzoekt. Op welke variabele(n) zou je je conclusies dan het liefst baseren?</p>  |   |   |  |
|     | <p>Om H3 te testen kunnen de conclusies ook om de hoeveelheid eten die geconsumeerd wordt gebaseren. Hierbij is het echter belangrijk om op het verschil tussen de consumptie van gezond eten en ongezond eten te letten.</p>   |   |   |  |
|     | <p><b>Geen antwoord</b></p> <p><i>of</i></p> <p><b>fout antwoord</b></p>  | <p><b>Antwoord</b></p> <p>“hoeveelheid” of “hoeveelheid voedsel” zonder erop in te gaan dat het verschil tussen gezond en ongezond eten belangrijk is</p> | <p><b>Antwoord</b></p> <p>hoeveelheid gezond eten of hoeveelheid ongezond eten, maar noemt niet zowel gezond als ook ongezond eten</p> <p><i>of</i></p> <p>Geeft aan dat gezond en ongezond eten worden vergeleken, omdat een (foute) conclusie is aangegeven</p> | <p><b>Antwoord dat men de hoeveelheid gezond en ongezond eten vergelijkt</b></p> |
| 3.4 | <p>Hoe wordt gemeten of de hoeveelheid voedsel tussen de condities verschilt?</p>   |   |   |  |
|     | <p>Camera's zijn opgesteld om het eetgedrag van de deelnemers eventueel nog achteraf te analyseren (r. 80-81)</p> <p>Een assistent telt voor- en achteraf het aantal voedsleenheden (dus de stukken) → om te kunnen berekenen hoeveel er per conditie gegeten was. Dit wordt met behulp van de videopnames gecontroleerd ( r. 87-90).</p> |   |   |  |
|     | <p><b>Geen antwoord</b></p> <p><i>of</i></p> <p><b>fout antwoord</b></p>  | <p><b>Camera's worden genoemd, geen verklaring hoe gemeten wordt dat</b></p>  | <p><b>In ieder geval staat beschreven dat het eten voor</b></p>   | <p><b>Er wordt zowel aangegeven dat het eten voor en</b></p>                     |

|     |   |   |   |   |
|-----|---|---|---|---|
|     |   | <b>er een verschil tussen de condities bestaat, alleen uitspraken die duidelijk maken dat de tekst juist weergegeven is (bv. “camera’s”, “tellen” of “assistent”)</b> | <b>en achteraf vergeleken wordt (bv. dat er getelt wordt hoeveel stukken er telkens in een bakje zitten)</b>  | <b>achteraf vergeleken wordt, als ook aangegeven dat camera’s gebruikt zijn om de resultaten te controleren</b> |
| 3.5 | Er is een probleem met de manier waarop de hoeveelheid voedsel die geconsumeerd is gemeten wordt. Weet jij welk probleem?   |   |   |   |
|     | Het probleem met deze methode van meten is, dat niet alle stukken gelijk groot en zwaar zijn. Dat betekent dat een gelijk aantal stukken, van gewicht duidelijk kan verschillen. Hierdoor zijn de resultaten niet betrouwbaar. De snelle muziek groep heeft duidelijk minder stukken gegeten, maar deze zijn misschien veel groter en vullen meer dan de stukken die de langzame muziek groep heeft gegeten. Gevulde koekjes vullen bv. veel meer dan druiven.  |   |   |   |
|     | <b>Geen antwoord of fout antwoord of antwoord die niet relevant is voor het probleem</b>  | <b>noemd het probleem dat wel relevant, maar niet essentieel is (bv. dat stukken verschillen vullen of dat dingen op de grond vallen)</b>                             | <b>In ieder geval duidelijk gemaakt dat meetmethode niet betrouwbaar is (bv. door duidelijk te maken verschillende hoeveelheden stukken hetzelfde gewicht kunnen hebben en hierdoor de meetresultaten niet met elkaar vergeleken kunnen worden)</b> | <b>Maakt duidelijk dat meetmethode niet betrouwbaar is door beide problemen aan te geven</b>                    |
| 3.6 | Kun je voor dit onderzoek een betere manier bedenken om te meten of de hoeveelheid voedsel die geconsumeerd wordt tussen de condities verschilt?  |   |   |   |
|     | Hier zijn verschillende antwoorden mogelijk. Met betrekking tot het hoofdprobleem, dat de hoeveelheid voedsel die geconsumeerd wordt niet betrouwbaar wordt gemeten is een mogelijke aanbeveling het gewicht van elk bakje voor en naar de gespreken te meten. Verder is het een mogelijk alleen voor voedsleenheden te kiezen die gelijk groot zijn en gelijk vullend. Het zal dus bijvoorbeeld een mogelijkheid zijn alleen een soort voedsel aan te bieden. Dit levert echter het probleem op, dat niet iedereen het aangeboden voedsel eet. |   |   |   |
|     | <i>Bij dit vraag is het belangrijk te bepalen of een idee met betrekking tot het onderzoek uitvoerbaar is en of het wellicht nieuwe problemen oplevert.</i>   |   |   |   |
|     | <b>Geen antwoord of fout antwoord</b>   | <b>Manier bedacht die in principe juist (en uitvoerbaar) is,</b>  | <b>Manier bedacht het probleem zou kunnen</b>   | <b>Juiste manier bedacht om het probleem</b>  |

|     |  |   |   |   |
|-----|--|---|---|---|
|     | <b>(die bv. met betrekking tot het onderzoek niet uitvoerbaar is)</b>  | <b>maar niets aan het probleem verbeterd</b>  | <b>verbeteren maar met betrekking tot het onderzoek niet echt uitvoerbaar is of nieuwe problemen oplevert (zoals bv. in het voorbeeld beschreven, alleen een soort voedsel te gebruiken)</b>  | <b>op te lossen, die uitvoerbaar is</b>   |
| 4.1 | <p>Kijk naar de resultaten in Tabel 1. Worden de resultaten uit de tekst op een correcte wijze weergegeven? Motiveer je antwoord.</p> <p>Nee, de resultaten uit de tekst worden niet op correcte wijze weergegeven.</p> <p>Probleem 1: In de controle groep hebben de personen in de tabel niet meer gegeten als in de snelle groep conditie, hetzelfde geldt voor de langzame muziek conditie. Verder werd volgens de tabel in de controle en de langzame conditie niet bijna even veel gegeten. De reden hiervoor is dat de snelle muziek conditie en de langzame muziek conditie omgedraaid zijn.</p> <p>Probleem 2: Verder verschillen in de tabel de arousal scores sterk van elkaar, hoewel in de tekst aangegeven is dat er geen verschillen in de arousal scores gevonden werden.</p> <p><i>Sommige deelnemers geven wellicht niet expliciet aan dat de tabel de tekst niet correct weergeeft. Als zij problemen noemen, betekent dit echter dat zij begrepen hebben dat er geen volledige overeenstemming is.</i></p> |   |   |   |
|     | <b>Geen antwoord of fout antwoord (die bv. vermeld dat de tabel correct is weergegeven)</b>  | <b>Alleen aangegeven dat de tekst en de tabel niet overeen komen, zonder een (correcte) verklaring wat er fout is</b> | <b>Aangegeven dat de tekst en de tabel niet overeen komen, maar alleen en van de beide problemen correct weergegeven (dus of het probleem dat de beide condities omgedraaid zijn of het probleem dat de arousal scores fout zijn weergegeven)</b> | <b>Aangegeven dat de tekst en de tabel niet overeenkomen en beide problemen correct weergegeven</b> |
| 4.2 | <p>Vergelijk de tabel en de grafiek. Valt je iets op?</p> <p>In vergelijking met de tabel is de grafiek fout. De staf van de snelle muziek conditie zou het hoogst moeten zijn. De staven van de langzame muziek</p>   |   |   |   |

|     |   |  |  |   |
|-----|---|--|--|---|
|     | conditie en de controle conditie zouden ongeveer gelijk hoog moeten zijn. De reden voor deze fouten is dat de controle muziek conditie en de snelle muziek conditie omgedraaid zijn.  |  |  |   |
|     | <b>Geen antwoord<br/>of<br/>fout antwoord<br/>(bv. dat de<br/>grafiek en de<br/>tabel niet<br/>verschillen)</b>   | <b>Alleen aangegeven<br/>dat de grafiek in<br/>vergelijking met de<br/>tabel niet correct is</b> | <b>Alleen<br/>aangegeven dat<br/>of snelle<br/>muziek of die<br/>controle<br/>conditie niet<br/>juist zijn<br/>weergegeven</b> | <b>Opgemerkt dat<br/>de staven van<br/>de controle<br/>conditie en de<br/>snelle muziek<br/>conditie<br/>uitgewisseld<br/>zijn<br/>of<br/>Zowel<br/>aangegeven,<br/>dat de staaf<br/>van de snelle<br/>muziek<br/>conditie het<br/>hoogst zou<br/>moeten zijn en<br/>dat die staven<br/>voor de<br/>langzame<br/>muziek<br/>conditie en de<br/>controle<br/>muziek<br/>conditie<br/>ongeveer even<br/>hoog zouden<br/>moeten zijn</b> |
| 4.3 | Stel dat Hypotheses 3 (H3: Het muziektempo beïnvloedt de hoeveelheid gezond voedsel die geconsumeerd wordt) ook onderzocht werd en de resultaten waren als volgt: In de snelle muziek conditie wordt significant meer gezond voedsel geconsumeerd dan in de controle conditie en de langzame muziek conditie. Er werd geen significant verschil tussen de controle conditie en de langzame muziek conditie gevonden. Schets hieronder hoe een grafiek uit zou kunnen zien. Gebruik de grafiek uit de tekst als voorbeeld. |  |  |   |

|  |  |  |   |  |
|--|--|--|---|--|
|  |  <p style="text-align: center;"><b>Conditie</b></p> <p><i>Alle staven juist de schetsen betekent in dit geval, dat de staaf voor de snelle muziek conditie hoger moet zijn dan de andere twee staven. Verder zullen de staven van de controle conditie en den langzame muziek conditie niet te duidelijk van elkaar verschillen of gelijk zijn. Voor een volledig correct plaatje is ook de correcte belettering van het plaatje van belang.</i></p>   |  |   |  |
| <p><b>Geen tekening of Fout tekening (bv. geen stafdiagram of geen belettering of alleen belettering)</b></p>  | <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td data-bbox="347 1198 619 2038"> <p><b>De snelle muziek staaf is het hoogst, maar het verschil tussen de staven van de controle conditie en de langzame muziek conditie is groter dan tussen de langzame en de snelle muziek conditie of tussen de controle en de snelle muziek conditie, er is een belettering of</b></p> <p><b>De staf van de snelle muziek conditie is niet het hoogst, maar het verschil tussen de staven van de controle conditie en de langzame</b></p> </td> <td data-bbox="619 1198 922 2038"> <p><b>Alle staven zijn blijkbaar juist weergegeven, maar er is geen belettering, de snelle muziek staaf is het hoogst en de staven van de controle conditie en de langzame muziek conditie verschillen minder sterk dan de staaf van de snelle muziek conditie en de langzame muziek conditie of de staaf van de snelle muziek conditie en de staaf van de controle</b></p> </td> <td data-bbox="922 1198 1418 2038"> <p><b>Alle staven zijn juist weergegeven, de grafiek is (correct) beletterd, de snelle muziek staaf is het hoogst en de staven van de controle conditie en de langzame muziek conditie verschillen minder sterk dan de staaf van de snelle muziek conditie en de langzame muziek conditie of de staaf van de</b></p> </td> </tr> </table> | <p><b>De snelle muziek staaf is het hoogst, maar het verschil tussen de staven van de controle conditie en de langzame muziek conditie is groter dan tussen de langzame en de snelle muziek conditie of tussen de controle en de snelle muziek conditie, er is een belettering of</b></p> <p><b>De staf van de snelle muziek conditie is niet het hoogst, maar het verschil tussen de staven van de controle conditie en de langzame</b></p> | <p><b>Alle staven zijn blijkbaar juist weergegeven, maar er is geen belettering, de snelle muziek staaf is het hoogst en de staven van de controle conditie en de langzame muziek conditie verschillen minder sterk dan de staaf van de snelle muziek conditie en de langzame muziek conditie of de staaf van de snelle muziek conditie en de staaf van de controle</b></p> | <p><b>Alle staven zijn juist weergegeven, de grafiek is (correct) beletterd, de snelle muziek staaf is het hoogst en de staven van de controle conditie en de langzame muziek conditie verschillen minder sterk dan de staaf van de snelle muziek conditie en de langzame muziek conditie of de staaf van de</b></p> |
| <p><b>De snelle muziek staaf is het hoogst, maar het verschil tussen de staven van de controle conditie en de langzame muziek conditie is groter dan tussen de langzame en de snelle muziek conditie of tussen de controle en de snelle muziek conditie, er is een belettering of</b></p> <p><b>De staf van de snelle muziek conditie is niet het hoogst, maar het verschil tussen de staven van de controle conditie en de langzame</b></p> | <p><b>Alle staven zijn blijkbaar juist weergegeven, maar er is geen belettering, de snelle muziek staaf is het hoogst en de staven van de controle conditie en de langzame muziek conditie verschillen minder sterk dan de staaf van de snelle muziek conditie en de langzame muziek conditie of de staaf van de snelle muziek conditie en de staaf van de controle</b></p>  | <p><b>Alle staven zijn juist weergegeven, de grafiek is (correct) beletterd, de snelle muziek staaf is het hoogst en de staven van de controle conditie en de langzame muziek conditie verschillen minder sterk dan de staaf van de snelle muziek conditie en de langzame muziek conditie of de staaf van de</b></p>   |   |  |

|     |  |  |  |   |
|-----|--|--|--|---|
|     |  | <b>muziek conditie is groter dan tussen de langzame en de snelle muziek conditie of tussen de controle en de snelle muziek conditie, er is een belettering</b>   | <b>conditie</b>  | <b>snelle muziek conditie en de staaf van de controle conditie</b>  |
| 5.1 | Vergelijk de conclusies en de resultaten. Benoem de conclusies die worden getrokken. Zijn deze correct? Motiveer je antwoord.  |  |  |   |
|     | De conclusies in de Discussie zijn als volgt:<br>Conclusie 1: bij langzame muziek wordt meer gegeten dan bij snelle muziek. (R. 135-136)<br>Conclusie 2: in de snelle muziek conditie wordt minder gegeten dan in de controle conditie. (R. 136-138)<br>Conclusie 3: geen duidelijk verschil tussen de controle en de langzame muziek conditie. (R. 138-140) |  |  |   |
|     | Alle drie Conclusies zijn met betrekking tot de tekst uit het resultaatendeel juist.   |  |  |   |
|     | <b>Geen antwoord of fout antwoord</b>  | <b>Een of twee conclusies juist aangegeven, maar niet of fout aangegeven dat alle conclusies juist zijn of alleen aangegeven dat alle conclusies juist zijn, zonder (correcte) redenering</b>  | <b>Alle conclusies juist opgelist, maar niet of fout aangegeven of de conclusies juist zijn of aangegeven dat de conclusies juist zijn en een of twee goede conclusies aangegeven</b>                          | <b>Alle conclusies juist aangegeven en aangegeven dat alle conclusies juist zijn</b>  |
| 5.2 | Welke conclusie ontbreekt er gezien de resultaten? Motiveer je antwoord.   |  |  |   |
|     | Gezien de resultatentekst ontbreekt er de conclusie dat geen verschil in arousal en/of plezier tussen de condities gevonden werd.  |  |  |   |
|     | <b>Geen antwoord of fout antwoord</b>  | <b>Alleen "arousal" en/of "plezier" als antwoord aangegeven, zonder de volledige (of correcte) conclusie aan te geven (dus bv. de conclusie dat er geen verschil in de hoeveelheid geconsumeerd voedsel bij verschillende niveaus van arousal en plezier</b> | <b>Antwoord dat of arousal OF plezier in de conclusiedeel mist, wel aangegeven hoe de volledige conclusie eruit ziet (dus bv. de conclusie dat er geen verschil in de hoeveelheid geconsumeerd voedsel bij</b> | <b>Antwoord dat de conclusie mist (dus bv. de conclusie dat er geen verschil in de hoeveelheid geconsumeerd voedsel bij verschillende niveaus van arousal en plezier gevonden werd)</b> |



|     |  | gevonden werd)  | verschillende niveaus van arousal gevonden werd)  |   |
|-----|--|---|---|---|
| 5.3 | Welk advies zou je een gewoon restaurant (dus geen "Eet-zoveel-als-je-wilt" restaurant) geven op basis van de resultaten van dit onderzoek? Motiveer je antwoord.  |   |   |   |
|     | In zo een restaurant zou men langzame of helemaal geen muziek moeten draaien, om te bereiken, dat de klanten zo veel mogelijk eten.  |   |   |   |
|     | <i>Het is mogelijk, dat deelnemers op dit vraag een antwoord geven die helemaal niet verwacht was. Als de antwoord desondanks op basis van wat in de verslag weergegeven wordt logisch is, krijgt de deelnemer een punt. Bv. als de deelnemer voorstellen voor bepaalde (langzame) muziek maken.</i> |   |   |   |
|     | <b>Geen antwoord of fout antwoord (die bv. op basis van de tekst niet logisch is)</b>  | <b>Fout antwoord die op basis van het verslag wel logisch is (bv. voorbeeld van een langzaam muziek stuk)</b> | <b>Alleen opgemerkt dat geen muziek gedraaid moet worden of alleen opgemerkt dat langzame muziek gedraaid moet worden</b> | <b>Aangegeven dat langzame of geen muziek gedraaid moet worden, omdat klanten dan meer eten</b> |