



## Gamebrics Project Deliverable:

### Resultaten evaluatiestudie

<b>Numer Deliverable</b>	D14
<b>Naam Deliverable</b>	Werkdocument met samenvatting belangrijkste bevindingen uit evaluatiestudie (D14)
<b>Werkpakket</b>	WP4: Effectmeting (uitvoering studie en evaluatie met twee games)
<b>Opleverdatum</b>	<b>Gepland:</b> 28-02-2023 <b>Actueel:</b> 18-10-2023
<b>Versie</b>	<b>Versie:</b> 1.0      Concept <input type="checkbox"/> Definitief <input checked="" type="checkbox"/>
<b>Auteurs</b>	Hans Hummel (OW)
<b>Verantwoordelijke/contactpers.</b>	Hans Hummel (OW)
<b>Kernwoorden</b>	Evaluatiestudie, onderzoekopzet, resultaten, leereffect, waardering

Gamebrics Project penvoerder: Open Universiteit Nederland  
Projectleiding en contactpersoon: Dr. Hans Hummel, [hans.hummel@ou.nl](mailto:hans.hummel@ou.nl)

Gamebrics deliverables mogen vrij gebruikt worden onder [CC-BY](https://creativecommons.org/licenses/by/4.0/) licentie (Creative Commons)

## Document revisie log

Version	Date	Description	Author
0.4	19-09-2023	Invoer en organisatie dataset in SPSS, nadat (met drie maanden vertraging) een volledige Excel sheet met data op 15/9 beschikbaar kwam vanuit Emergo. Eerste analyses en resultaten gescand.	Hans Hummel (OW)
0.9	22-09-2023	Volledige conceptversie ter informatie en goedkeuring laatste PT op 17-10-2023	Hans Hummel (OW)
1.0	18-10-2023	Vastgesteld op PT van 17-10-2023, laatste redactie	Hans Hummel (OW)



## TABLE OF CONTENTS

Summary.....	4
1. Research setup .....	5
1.1. Participants .....	6
1.2. Procedure .....	7
1.3. Instruments .....	8
2. Main performance results .....	10
3. Main perception results .....	13
4. Conclusions and recommendations .....	17
Appendices Gamebrics questionnaires	
Appendix 1.1. Pre/post-test knowledge questions Kastanjehoeve .....	19
Appendix 1.2. Pre/post-test knowledge questions Junior scientist .....	25
Appendix 2. (Sub)scales and appreciation questionnaire .....	32

## Summary

This document contains main results from the evaluation study we designed and executed within the context of the Gamebrics project, which is co-financed by the Dutch Ministry of education under the Surf call 'Open Higher Education 2020' (project number OO21-01). This project designs, develops and evaluates analytical rubrics integrated in the gameplay of serious games, as formative assessment support when acquiring professional competences.

This document is about *the evaluation part*, and describes the research setup, data collection and data analysis of an evaluation study carried out with two serious games in educational practice. Data were collected from computer logging and questionnaires (embedded in the game-play environment), and were of both subjective (perception scores) and objective (performance) nature.

This text is written in English because it not only used for deliverable 14 (D14), called "Workdocument containing most important findings from evaluation study", but also for an academic article (to be submitted as deliverable 19.2) in an international, peer-reviewed and high impact factored journal in the field of Educational Technology.

In this deliverable we describe the research setup (section 1), without repeating too much detail (can be found in Deliverable 10/12). This Deliverable 14 will focus on final procedures followed (including the final population that participated and instruments used) and main findings found. We will describe main performance results (section 2) and main perception results (section 3). We will end with a conclusion and some recommendations (section 4).

In summary, our data analyses could be carried out with the datasets of a total of eighty-four participants (two thirds studying one, one third studying the other game), and we found some promising results that support the approach taken in the project.

When it comes to subjective *student appreciations* of feedback provided (both 'natural feedback' as part of the game scenario as the 'reflective' feedback in the Gamebrics dashboard), it appeared "satisfactory". We questioned aspects like perceived pressure and competence (regarding motivation), as well as appreciation, understanding and need for the feedback provided (regarding the perception of feedback). It is interesting to see that students with reflective feedback (on competence) perceived their competence (development) better. These findings confirm the second research question (Do students with the support of the Gamebrics dashboard appreciate this kind of feedback?).

When it comes to *learning effects*, we first of all found that all students do seem to learn from playing the games, an effect that was stronger when playing the games for a longer time, and an effect that was more evident with one than with the other game. Overall, the students that received dashboard information show higher learning increases (when comparing their post- and pre-test scores on knowledge), and also more efficient gameplay behavior (when looking at in-play performance scores) when compared to students that did not receive the Gamebrics tooling. These students in the experimental condition also showed more efficient in-game behavior (when looking at their performance scores). These findings confirm the second research question (Do students with the support of the Gamebrics dashboard learn more effectively?).

## 1. Research setup, procedure and research questions

This section describes research setup, prior procedures and research questions.

### RESEARCH SETUP

To research the effects of the Gamebrics dashboard we randomly allocated half of participants in an experimental group (with dashboard) and another half in a control group (without dashboard). We designed two measurement instruments: (a). the pre-/post-test knowledge test contains 2x20 awareness questions administered before and after gameplay; and (b). an appreciation questionnaire (with either 27 or 43 items on motivation and feedback) with appreciation questions to be administered after gameplay (to answer RQ2: Do learners appreciate the dashboard information?). RQ1 will be answered by comparing pre- with post-test scores, in combination with data (time, performance) that will be obtained from Computer logging data.

We first checked for direction of items and normality of score distribution. Besides ANOVA's as main test for expected effects (and differences between conditions), we also applied other statistics. A paired t-test was calculated to determine (significant) learning effects by gameplay, overall and for both games. Statistical descriptives (means, standard deviations) were calculated for scoring of (sub)scales of the appreciation questionnaire (and compared both within groups and between games). Participants (when having consented to participate, which was the case in 73% in the student population) were given access to the game contained in their course and randomly allocated to the control group (standard game with just 'natural feedback') or the experimental group (game with additional Gamebrics dashboard for 'reflective feedback'), see Figure 1.

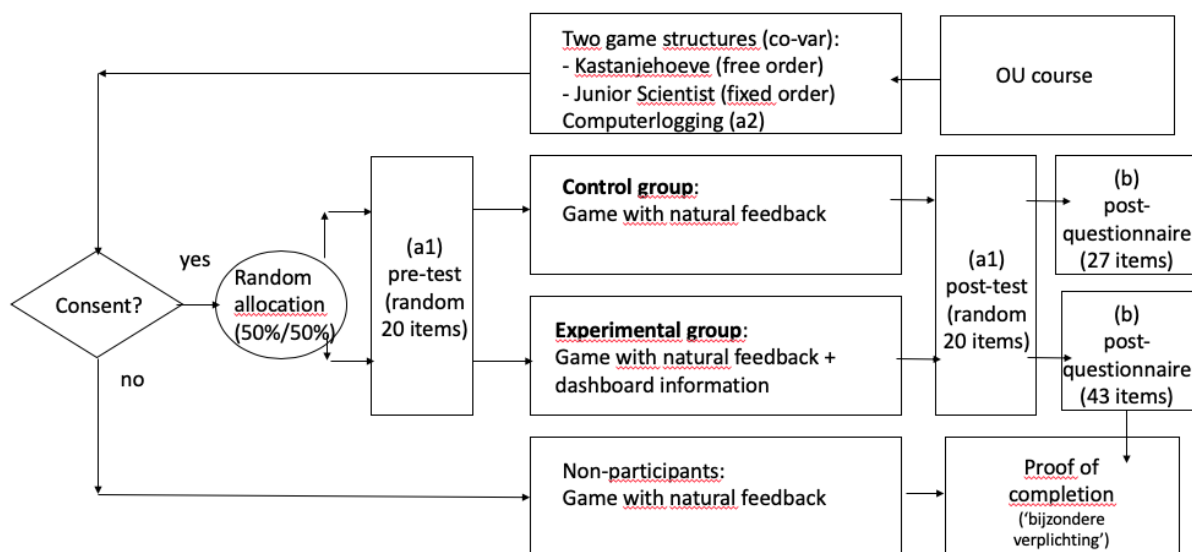


Figure 1. Setup Gamebrics study

## *PRIOR PROCEDURES*

The CETO procedure (approval to carry out the study from OU's Ethical Research Committee) was started by writing and sending an elaborate application form (information also available in combined D10 and D12), both delivered on May 31th, 2022. The approval from CETO was then received on July 11th, 2022, after having provided some additional information on their request.

The research setup (see Figure 1) has been described in some more detail in project deliverable D10/12, and we will therefore focus on the main findings in this deliverable (repeating some detail in the appendices).

The elaborated questionnaire items could already be included in the CETO application, now see appendix 2. The elaborated pre/post-test questions became available later, and were then added to the D10/D12 deliverable, on July 8, 2022 (for Kastanjehoeve game) and November 4, 2022 (for Junior Scientist game) respectively, see appendices 1.1 and 1.2.

Both games with integrated Gamebrics tooling became available and were launched for the evaluation study at two different moments, early September 2022 (Kastanjehoeve game) and early February 2023 (Junior Scientist game) respectively, due to unexpected delays in game development.

## *RESEARCH QUESTIONS*

When we launched the evaluation environments, half of participants (experimental groups) were and half (control groups) were not provided with the additional dashboard information in between mini-games. Most of our serious games consist of 'mini-games' or challenges; these are self-contained entities with some 'closure' or intermediate gameplay result. Games under study were not only different by topic, but also by structure and complexity level. Participants that completed the Kastanjehoeve game (starter course on Marketing & Organisation) solved some more loosely connected problems. Participants that completed the Junior Scientist game (advanced course on Experimental research in Psychological) study solved more complex and connected problems.

However, for both games under study, the overall purpose of the empirical study is to assess if such (EMERGO) games that contain additional Gamebrics tooling (progress on an analytical rubric depicted in a dashboard) are more effective and satisfactory in helping students acquire analytical (problem solving) skills through active gameplay. Main research questions therefore are:

RQ1. Do students with the support of the feedback dashboard learn more effectively?; and

RQ2. Do students with the support of the feedback dashboard appreciate this kind of feedback?

## **1.1. Participants**

Complete data sets could finally be collected from eighty-four participants that agreed to participate (after receiving an information letter and by informed consent) in the study, and then had completed all game play (and thus data collection). Fifty-six of them studied the Kastanjehoeve game (Management and Organisation) and twenty-eight studied the Junior Scientist game (Psychology).

Participants were randomly allocated to game condition (without or with Gamebrics tooling, control group or experimental). Overall (that is for the total population), half (forty-two students) were indeed in the experimental group, and half (forty-two students) in the control group. For the Kastanjehoeve game, this was twenty-seven students in the control group and twenty-nine in the experimental group.



For the Junior Scientist game, this was fifteen students in the control group and thirteen in the experimental group.

In the population we had thirty-six students (42,9%) of male gender and forty-eighth (57,1%) of female gender. Their average age was 32,64 years, ranging from 18 to 65 ( $SD = 12,328$ ).

## 1.2. Procedure (within evaluation period)

The period of data collection turned out much longer than originally envisioned when writing the project proposal. First of all, the development of the Junior Scientist game (and Gamebrics tooling) game became delayed due to non-sufficient development capacity at the ECO department of the OUNL, where all (remaining) EMERGO development capacity had been placed since 2020. Junior Scientist could therefore not be launched for data collection together with Kastanjehoeve in September 2022, but not until February 2023.

When we started data collection, about 260 students were enrolled for both courses and gameplay (a lower number than the initially expected 400 students at time of proposal writing). However, about 27% of the students that started playing the games did not want to participate in the evaluation study. Even more important, most students appeared to need much time to arrive at the end and complete game play (and data collection). Finally, it appeared more troublesome to extract the collected data from EMERGO, and deliver it as an integrated dataset. As a consequence of these factors, the period of data collection (originally planned till May 2023) was extended till September 2023.

When transferring the data to SPSS (version 24) for further analysis, the first steps were to control for normality of distribution and for (same) direction of items in the appreciation questionnaire.

Table 1. Check for Normality of distribution main Likert-scores (skewness and kurtosis)

	N	Mean	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
PERFORMANCE	84	,818841	-1,151	,263	,895	,520
MOT_PRES_AVG	84	3,5667	-,463	,263	,870	,520
MOT_COMP_AVG	84	3,0159	-,488	,263	,727	,520
FB_ACC_AVG	84	3,3175	-,743	,263	1,147	,520
FB_UND_AVG	84	3,2177	-,182	,263	,674	,520
FB_NEED_AVG	84	3,3532	,214	,263	-,434	,520
Valid N (listwise)	84					

Normality of appreciation subscales was assessed on the bases of skewness and kurtosis scores (see Table 1), and it was concluded that the data are approximately normally distributed, based on

Cameron's (2004, p. 543) argument that "skewness and kurtosis should both fall in the range from +2 to -2 if data are normally distributed".

The forty-three items of the appreciation questionnaire were checked on having the same 'direction'. In order for all 5-point Likert items to go in the same direction, that is from 1 score (most negative) to 5 scores (most positive), twenty out of forty-three had to be recoded, that is: 1 became 5, 2 became 4, 4 became 2, and 5 became 1 scores. These items were numbers 2, 4, 5, 11, 13, 15, 19, 21, 23, 25, 26, 27, 29, 31, 35, 37, 39, 41, 42 and 43 (see appendix 2).

### 1.3. Instruments

#### *PRE/POST TEST*

We asked our (experienced) teachers / project members to draw up a set of (about forty) awareness questions in MC format (about the content of the respective game, that is authentic cases on either human resource management or on experimental research design). Two times forty questions were used in a pre-/post-test, with twenty pre- and twenty post-test items, see appendices 1. Both pre- and post-test questions were presented in random order to each participant, also to further make harder potential study fraud attempts. Later upon completion of the game and for purposes of data analysis (comparison), they were extracted to datasets in the same order.

As you can see in the appendices 1, every (closed) question has five answer options, of which only one is right (r) and four being false (f). Obviously, what were right/false answers was not visible to the students. Half of the randomized subset of twenty items (automatized by EMERGO) was provided before starting gameplay; the other half of twenty items was provided at the end of gameplay. Deltas between (average) scores after and before show growth on knowledge/awareness. A between- groups comparison of (average) scores on the pre-test checked if baselines were about equal for both conditions, which was the case. At the start we found mean baseline scores of 9,96 (control,  $N = 27$ ,  $SD = 2,738$ ) and 10,10 (experimental,  $N = 29$ ,  $SD = 2,920$ ) for Kastanjehoeve, and mean scores of 12,00 (control,  $N = 15$ ,  $SD = 2,268$ ) and 11,92 (experimental,  $N = 13$ ,  $SD = 2,253$ ) for Junior Scientist. In both cases, such a small difference of about 0,1 point (only 0,5% of maximum score to be achieved) can be neglected without carrying out further statistical descriptives.

We had already successfully used this pre-/post-test approach to examine (and proof) significant learning growth through gameplay in two previous (and published) studies (Hummel et al., 2020; 2021), one in an academic context (university course on Psychology) and the other in a vocational training context (VET training on entrepreneurship in construction). When we tested the internal consistency of these questions in both studies, we found overall (Cronbach's) Alphas of  $\alpha = .656$  and  $\alpha = .696$  respectively for internal consistency, which is considered 'acceptable' (e.g., Pallant, 2001). For this study we found similar Alphas for the internal consistency between the forty pre/post test questions, respectively  $\alpha = .654$  for the questions about the content of the Kastanjehoeve game and an even somewhat higher  $\alpha = .774$  for Junior Scientist questions.

#### *APPRECIATION QUESTIONNAIRE*

For appreciation of gameplay features, we had already used (validated) questionnaire items with the same game feature scales to measure perceived motivation, each item using a five-point Likert scale (with values from 1 = completely disagree to 5 = completely agree), based on (validated) scales (Nadolski & Hummel, 2017; Hummel et al., 2020).





In this study we were also interested in the perceived appreciation of feedback / formative assessment within the gameplay (and not gameplay as such), and we developed specific subscales on (perceived) 'acceptance', 'understanding' and 'need' for such feedback. These new subscales had not previously been checked for internal consistency and validity.

This questionnaire (see appendix 2 for subscales and items) was also administered by the EMERGO environment upon completion of gameplay; and all collected answers were exported to an Excel file for further analysis.

For measuring perceived motivation during gameplay, this study used the existing IMI-subscales 'perceived pressure' (5 items,  $\alpha = 0,828$  for this study) and 'perceived competence' (6 items,  $\alpha = 0,825$  for this study), which had been validated in numerous studies.

For measuring the appreciation of the feedback provided (which is in the core interest of this study), this study used specific and new subscales for measuring 'feedback acceptance' (subscale 1, with 6 items,  $\alpha = 0,758$  for natural and  $\alpha = 0,771$  for reflective feedback), 'feedback understanding' (subscale 2, with 7 items,  $\alpha = 0,775$  for natural and  $\alpha = 0,884$  for reflective feedback), and 'feedback need' (subscale 3, with 3 items,  $\alpha = 0,328$  for natural and  $\alpha = 0,590$  for reflective feedback). These internal consistency measures can be considered 'acceptable' for all subscales, except for the 'feedback need' subscale, which has to do with too limited amount (three) of items.

For the feedback subscales we had to differentiate between 'natural feedback' (control and experimental group) and 'reflective feedback' provided by the dashboard (only experimental group). All feedback items therefore had two variants, meaning that the questionnaire for the experimental group was longer. For participants in the control group, the total questionnaire contained 27 items (11+6+7+3). For participants in the experimental group, the total questionnaire contained 43 items (11+(2\*(6+7+3) = 11 + 32), see Appendix 2. Items 28-33, items 34-40, and items 41-43 will only be presented to and answered by participants in the experimental group.

#### *COMPUTER LOGGING*

A dedicated research data component of the EMERGO authoring environment logged and extracted various gameplay activities as potentially interesting dependent variables. For this study we have looked into total playtime (computer clock time converted to minutes), total time for studying the dashboard, and in-game performance scores. This is because we want to know how much (additional) time students (in the experimental condition) spend on consulting the (additional) tooling (the dashboard information), and if studying this tooling has any effect on the in-game performance.

This performance score is not only based on the total scores obtained (product scores), but also on monitoring all errors made during gameplay and later transformed into 0-100% scores, with a 100% score indicating that no errors were made by student (most efficient learning) and a 0% score indicating 'random behavior' (or less) by student (least efficient learning). These (process) scores depend on the number of alternatives (nAlt) and amount of errors (nErr) for each activity. All game activities are monitored, including answering MC questions, multi-select tasks, opening and closing resources, categorizing items (using drag-and-drop tools), selecting text fragments, and others.

## 2. Main performance results

First of all, we wanted to check if playing the games had any learning effect at all on knowledge gain, apart from being fun to do, and still independent of a (potential) contribution of the additional Gamebrics tooling. The learning effect for both games together (that is not differentiating for individual game effects) appeared to be quite high, when comparing all post-test scores with all pre-test scores (learning growth), with  $t(1, 84, \text{paired}) = 4,664$ ,  $p < 0,001$ . The average growth for both games was 1,75 point (8,75% of maximum score to be achieved), going from a 10,68 to 12,43 score, see Table 2.

Table 2. Average Pre/Post–test scores overall for both games together

Both games		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	POSTTEST	12,43	84	3,298	,360
	PRETEST	10,68	84	2,769	,302

However, when differentiating for both games, the results are quite diverse.

An *overall* learning effect for the Kastanjehoeve game (independent of condition) seems to be even more substantial, with  $t(1, 56, \text{paired}) = 5,864$ ,  $p < 0,001$ . The average growth found for this game was 2,45 point (12,25% of maximum score to be achieved), going from a 10,04 to 12,48 score, see Table 3.

Table 3. Average Pre/Post–test scores overall for Kastanjehoeve

Kastanjehoeve		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	POSTTEST	12,48	56	2,835	,379
	PRETEST	10,04	56	2,809	,375

On the contrary, an *overall* learning effect for the Junior Scientist game (independent of condition) could *not* be found, with  $t(1,27, \text{paired}) = 0,514$ ,  $p = 0,611$ . The average growth found for all students that played this game was only 0,36 point (1,8% of maximum score to be achieved), going from a 11,96 to 12,32 score, see Table 4.

Table 4. Average Pre/Post–test scores overall for Junior Scientist

Junior Scientist		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	POSTTEST	12,32	28	4,128	,780
	PRETEST	11,96	28	2,219	,419

When we also differentiate for condition in both games, we better come to understand this striking difference in the overall effect on growth found between games.

For this, we have to focus on the average pre- and post-test scores for both conditions of the Junior Scientist game, see Table 5. It turns out that the average growth for students in the control group playing Junior Scientist (with  $N = 15$ ) was not only missing, but even negative ( $M = -0,80$ ,  $SD = 4,459$ ). Students in the experimental group playing Junior Scientist (with  $N = 13$ ) did show a significant (positive) growth ( $M = 1,69$ ,  $SD = 1,888$ ).



Table 5. Average Pre/Post–test scores for both conditions Junior Scientist

Junior Scientist		PRETEST	POSTTEST
Control (without dashboard)	Mean	12,00	11,20
	N	15	15
	Std. Deviation	2,268	5,060
Experimental (with dashboard)	Mean	11,92	13,62
	N	13	13
	Std. Deviation	2,253	2,256
Total	Mean	11,96	12,32
	N	28	28
	Std. Deviation	2,219	4,128

This explains why condition also does have a significant effect on learning increase for this Junior Scientist game, but that at the same time we could *not* find an *overall* learning effect for playing this game. A probable explanation might be this game and / or the post-test is of a higher complexity level, which is also reflected in lower in-game performance scores (see later).

For the Kastanjehoeve game, the findings appear more consistent. Students in both conditions show significant learning growth of  $M = 1,67$ ,  $SD = 2,882$  (with  $N = 27$ ) and  $M = 3,17$ ,  $SD = 3,208$  (with  $N = 29$ ) respectively, which comparison between games *and* conditions is reflected in Table 6.

Table 6. Average growth for both conditions, comparing both games

**Descriptive Statistics**

Dependent Variable: GROWTH

COND	GAME	Mean	Std. Deviation	N
Control (without dashboard)	Kastanjehoeve	1,67	2,882	27
	Junior Scientist	-,80	4,459	15
	Total	,79	3,673	42
Experimental (with dashboard)	Kastanjehoeve	3,17	3,208	29
	Junior Scientist	1,69	1,888	13
	Total	2,71	2,924	42
Total	Kastanjehoeve	2,45	3,121	56
	Junior Scientist	,36	3,674	28
	Total	1,75	3,439	84

Not only did we find significant effects for condition on learning increase (beyond gameplay), we also found a significant effect of condition on in-game performance. Furthermore, the amount of time spent on gameplay (in both conditions), and time studying the dashboard (experimental condition) is quite different for both games. Students spent three times more time on the Kastanjehoeve game (M

= 514, SD = 300) than on the Junior Scientist game (M = 175, SD = 80); which is about 8,5 hours compared to 3 hours, both with a rather large variance in times reported.

Both condition, game and time seem to be (cor)related to learning increase and in-game performance, as Tables 7 (Means) and 8 (Correlations) show us. However, a combined condition \* game effect could not be found in ANCOVA, for instance for growth, with  $F(1,80) = 0,439$ ,  $p = 0,510$ , partial Eta = 0,005.

Table 7. Average growth, performance and gameplay time for both games

		N	Mean	Std. Deviation	Minimum	Maximum
GROWTH	Kastanjehoeve	56	2,45	3,121	-4	9
	Junior Scientist	28	,36	3,674	-10	5
	Total	84	1,75	3,439	-10	9
PERFORMANCE	Kastanjehoeve	56	,860778	,0471841	,7018	,9286
	Junior Scientist	28	,734968	,0841465	,5634	,8694
	Total	84	,818841	,0856643	,5634	,9286
TIME_GAME	Kastanjehoeve	56	541,09	300,257	119	1363
	Junior Scientist	28	175,43	79,892	63	367
	Total	84	419,20	303,130	63	1363

Table 8. Correlations between gameplay time, time spent studying the dashboard, growth and performance

		Correlations			
		TIME_GAME	TIME_DASH	GROWTH	PERFORMANCE
TIME_GAME	Pearson Correlation	1	,172	,221*	,381**
	Sig. (2-tailed)		,118	,043	,000
	N	84	84	84	84
TIME_DASH	Pearson Correlation	,172	1	,248*	,319**
	Sig. (2-tailed)	,118		,023	,003
	N	84	84	84	84
GROWTH	Pearson Correlation	,221*	,248*	1	,283**
	Sig. (2-tailed)	,043	,023		,009
	N	84	84	84	84
PERFORMANCE	Pearson Correlation	,381**	,319**	,283**	1
	Sig. (2-tailed)	,000	,003	,009	
	N	84	84	84	84

\*. Correlation is significant at the 0.05 level (2-tailed).

\*\*. Correlation is significant at the 0.01 level (2-tailed).

Effects of condition on learning growth ( $F(1, 82) = 7,422$ ,  $p < .01$ ), in-game performance ( $F(1, 82) = 77,250$ ,  $p < .001$ ), and game time ( $F(1, 82) = 39,889$ ,  $p < .001$ ) are confirmed by ANOVA, see Table 9. The last finding can probably be explained by extra time spent on studying the dashboard (in experimental groups), although this additional time appeared relatively low (about 4% extra time).

Most important are the first two main findings / effects that provide us with a positive answer on the first RQ1: YES, offering the Gamebrics tooling / dashboard helps students learn more effective



Table 9. Correlations between gameplay

ANOVA		df	F	Sig.
GROWTH	Between Groups	1	7,422	,008
	Within Groups	82		
	Total	83		
PERFORMANCE	Between Groups	1	77,250	,000
	Within Groups	82		
	Total	83		
TIME_GAME	Between Groups	1	39,889	,000
	Within Groups	82		
	Total	83		

So, although 4% extra time spent on the dashboard (in both games) is not enough to speak of a time-on-task effect, it appears to account for causing (significantly) more time on gameplay (5 and 7,5% in both games respectively). This can also be considered to be a nice side-effect, especially in combination with more learning growth and better in-game performance scores.

### 3. Main perception results

When looking for findings on more subjective (or perception) scores, in order to answer the second RQ2 (Do students with the support of the feedback dashboard appreciate this kind of feedback?), we now first will look at average subscale scores. These scores pertain to students' appreciation of their motivation during gameplay and of the provided feedback when playing (that is as perceived by students after having played the game), feedback both as offered within the game scenario (natural feedback) and when studying the dashboard (reflective feedback), see Table 10. All average perception scores range between 3,0 and 3,6 (so between 'neutral' and 'agree'), and overall can be considered 'satisfactory'.

Table 10. Means for time, performance, growth and appreciation measures (8 subscales),

	Descriptive Statistics				
	N	Minimum	Maximum	Mean	Std. Deviation
TIME_TOT (min)	84	63	1370	427,37	306,069
TIME_DASH (min)	84	0	85	8,17	15,025
TIME_GAME (min)	84	63	1363	419,20	303,130
PERFORMANCE (0-1)	84	,5634	,9286	,818841	,0856643
PRETEST (0-20)	84	3	17	10,68	2,769
POSTTEST (0-20)	84	0	18	12,43	3,298
GROWTH	84	-10	9	1,75	3,439

MOT_PRES_AVG	84	1,00	5,00	3,5667	,81008
MOT_COMP_AVG	84	1,33	4,17	3,0159	,56541
FB_ACC_AVG	84	1,00	4,67	3,3175	,61367
FB_ACC_DASH_AVG	42	1,17	4,50	3,1944	,62134
FB_UND_AVG	84	1,29	4,86	3,2177	,63867
FB_UND_DASH_AVG	42	1,29	4,43	3,0408	,73880
FB_NEED_AVG	84	2,00	4,67	3,3532	,60862
FB_NEED_DASH_AVG	42	1,00	5,00	3,2063	,76846
Valid N (listwise)	42				

Table 11. Means for most important dependent variables per condition

		N	Mean	Std. Deviation	Minimum	Maximum
GROWTH	Control (without dashboard)	42	,79	3,673	-10	6
	Experimental (with dashboard)	42	2,71	2,924	-4	9
	Total	84	1,75	3,439	-10	9
PERFORMANCE	Control (without dashboard)	42	,791688	,0925129	,5634	,9107
	Experimental (with dashboard)	42	,845995	,0691843	,6231	,9286
	Total	84	,818841	,0856643	,5634	,9286
MOT_PRES_AVG	Control (without dashboard)	42	3,5286	,82618	1,00	5,00
	Experimental (with dashboard)	42	3,6048	,80181	1,00	5,00
	Total	84	3,5667	,81008	1,00	5,00
MOT_COMP_AVG	Control (without dashboard)	42	2,8373	,65383	1,33	4,17
	Experimental (with dashboard)	42	3,1944	,39289	2,67	4,00
	Total	84	3,0159	,56541	1,33	4,17
FB_ACC_AVG	Control (without dashboard)	42	3,3849	,58607	2,33	4,67
	Experimental (with dashboard)	42	3,2500	,63996	1,00	4,33
	Total	84	3,3175	,61367	1,00	4,67
FB_UND_AVG	Control (without dashboard)	42	3,2891	,60895	2,00	4,86
	Experimental (with dashboard)	42	3,1463	,66670	1,29	4,43
	Total	84	3,2177	,63867	1,29	4,86

	Total	84	3,2177	,63867	1,29	4,86
FB_NEED_AVG	Control (without dashboard)	42	3,3810	,64841	2,00	4,67
	Experimental (with dashboard)	42	3,3254	,57258	2,00	4,67
	Total	84	3,3532	,60862	2,00	4,67

When differentiating for condition, the range of scores becomes somewhat broader on the low end (from 2,8 to 3,6), with ‘perceived competence’ being the only somewhat ‘non-satisfactory’ aspect for the control condition, see Table 11. We also see that having the dashboard has a positive effect (only) on that specific perception and not on any other, see ANOVA Table 12. In other words: seeing information on your progress on certain subskills (competences) makes you feel more confident.

Table 12. *Effects of condition on perceptions*

ANOVA		df	F	Sig.
MOT_PRES_AVG	Between Groups	1	,184	,669
	Within Groups	82		
	Total	83		
MOT_COMP_AVG	Between Groups	1	9,207	,003
	Within Groups	82		
	Total	83		
FB_ACC_AVG	Between Groups	1	1,015	,317
	Within Groups	82		
	Total	83		
FB_UND_AVG	Between Groups	1	1,051	,308
	Within Groups	82		
	Total	83		
FB_NEED_AVG	Between Groups	1	,173	,678
	Within Groups	82		
	Total	83		

Finally, the correlation matrix relating all subscale scores for appreciation of motivation and feedback shows all subscales to be related, again with the exception of perceived competence with need for natural feedback, see Table 13. The acceptance of, understanding of and need for natural and reflective feedback seem to be related, see Table 14.

Table 13. Correlations between subscales natural feedback

		Correlations				
		MOT_PRES_	MOT_COMP_	FB_ACC	FB_UND_	FB_NEED_
		AVG	AVG	_AVG	AVG	AVG
MOT_PRES_AVG	Pearson Correlation	1	,358**	,429**	,410**	,252*
	Sig. (2-tailed)		,001	,000	,000	,021
	N	84	84	84	84	84
MOT_COMP_AVG	Pearson Correlation	,358**	1	,279*	,272*	,089
	Sig. (2-tailed)	,001		,010	,012	,423
	N	84	84	84	84	84
FB_ACC_AVG	Pearson Correlation	,429**	,279*	1	,784**	,538**
	Sig. (2-tailed)	,000	,010		,000	,000
	N	84	84	84	84	84
FB_UND_AVG	Pearson Correlation	,410**	,272*	,784**	1	,576**
	Sig. (2-tailed)	,000	,012	,000		,000
	N	84	84	84	84	84
FB_NEED_AVG	Pearson Correlation	,252*	,089	,538**	,576**	1
	Sig. (2-tailed)	,021	,423	,000	,000	
	N	84	84	84	84	84

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

Table 14. Correlations between subscales reflective feedback

		Correlations		
		FB_ACC_DASH_	FB_UND_DASH_	FB_NEED_DASH_
		AVG	AVG	AVG
FB_ACC_DASH_AVG	Pearson Correlation	1	,794**	,581**
	Sig. (2-tailed)		,000	,000
	N	42	42	42
FB_UND_DASH_AVG	Pearson Correlation	,794**	1	,603**
	Sig. (2-tailed)	,000		,000
	N	42	42	42
FB_NEED_DASH_AVG	Pearson Correlation	,581**	,603**	1
	Sig. (2-tailed)	,000	,000	
	N	42	42	42

\*\* . Correlation is significant at the 0.01 level (2-tailed).

When looking for findings on more subjective (or perception) scores in order to answer the second RQ2 (Do students with the support of the feedback dashboard appreciate this kind of feedback?), we found on overall 'satisfactory' appreciation on all eight aspects. It was especially noteworthy to see that having reflective feedback help student feel better about their competence. This second question can therefore also be answered positive.





#### 4. Conclusion and recommendations

Our data analyses could be carried out with the datasets of a total of eighty-four participants (two thirds studying one game, one third studying the other game), and we found some promising empirical results that support the approach taken in the Gamebrics project.

When it comes to subjective *student appreciations* of feedback provided (both ‘natural feedback’ as part of the game scenario as the ‘reflective’ feedback in the Gamebrics dashboard), the overall student appreciation appeared “satisfactory”. We questioned aspects like perceived pressure and competence (regarding motivation), as well as appreciation, understanding and need for the feedback provided (regarding the perception of feedback). It is interesting to see that students with reflective feedback (on competence) perceived their competence (development) better. These findings confirm the second research question (Do students with the support of the Gamebrics dashboard appreciate this kind of feedback?).

When it comes to objective *learning effects*, we first of all found that all students do seem to learn from playing the games, an effect that was stronger when playing the games for a longer time, and an effect that was more evident with one game than with the other game. Overall, the students that received dashboard information show higher learning increases (when comparing their post- and pre-test scores on knowledge), and also show more efficient in-game gameplay behavior (when looking at in-play performance scores) when compared to students that did not receive the Gamebrics tooling. These findings confirm the second research question (Do students with the support of the Gamebrics dashboard learn more effectively?).

Furthermore, we could observe that students did not suffer performance or usability problems with studying the dashboard and / or playing the games and /or the question items, after the Gamebrics tooling and questionnaires had been integrated in the gaming environment.

There are some limitations to be mentioned. During this evaluation study we could not research the usability of the Gamebrics Authoring environment, allowing teachers / developers to enter the rubric information and feedback texts, since this input tool was not yet available. Furthermore, such a usability test would not have been valid, since the two games used for the study were (largely) ready games with already existing game activities. This tool will become available at the end of the project. Beyond the scope, it would be recommendable to further test that tool when teachers / developers have to design a game with rubrics from scratch.

Another limitation is that the main findings reported in this deliverable are analyzed at the level of main dependent variables (scales or subscales), and not on concrete item level. To quantitatively estimate the success of the Gamebrics approach, this has been sufficient. For closer inspection of the game content and natural feedback provided with the scenario, it would be recommendable to look at the level of specific micro-challenges and items of subscales. It might well be that performance scores differ between micro-challenges, and / or that some items are scored different. We already saw large variances on main (sub) scales.

## REFERENCES

- Cameron, C. A. (2004). Kurtosis. In M. Lewis-Beck, A. E. Bryman, & T. F. Liao (Eds.), *The SAGE Encyclopedia of Social Science Research Methods* (pp. 543-544). London: SAGE Publications. doi:10.4135/9781412950589.
- Fang, X., Zhang, J., & Chan, S. (2013). Development of an instrument for studying flow in computer game play. *International Journal of Human-Computer Interaction, 29*, 456–470.
- Hummel, H. G. K., Nadolski, R. J., Eshuis, J., Slootmaker, A., & Storm, J. (2020). Serious game in introductory psychology for professional awareness: Optimal learner control and authenticity. *British Journal of Educational Technology, 52*(1), 125-141.
- Hummel, H. G. K., Slootmaker, A., & Storm, J. (2021). Mini-games for entrepreneurship in construction: Instructional Design and effects of the TYCON game. *Interactive Learning Environments*. DOI: 10.1080/10494820.2021.1995759
- Lewis, J. R., Utesch, B. S., & Maher, D. E. (2015). Measuring perceived usability: The SUS, UMUX-LITE, and AltUsability. *International Journal of Human-Computer Interaction, 31*, 496–505.
- Nadolski, R. J., & Hummel, H. G. K. (2017). Retrospective cognitive feedback for progress monitoring in serious games. *British Journal of Educational Technology, 48* (6), 1368-1379.
- Pallant, J. (2001), *SPSS survival manual - a step by step guide to data analysis using SPSS for windows* (version 10), Buckingham Open University Press.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78.

The anonymized dataset used for the analyses reported has been made publicly available through KNAW's DANS EASY repository of datasets at: <https://doi.org/10.17026/dans-zfe-p7ty>



## Appendix 1.1. Pre/post test vragen Kastanjehoeve

### Nota Bene

Per game zijn alle 40 items MC vragen met vier antwoordalternatieven waarvan één correct (aangegeven met vet) worden.

Bij elk item hieronder staat per game aangegeven over welke mini-game daarbinnen deze gaat, maar dat krijgt student NIET te zien. Items worden random uit deze set van 40 getrokken en in willekeurige volgorde en nummering aangeboden; eerste steekproef van 20 items als pre-test en overgebleven steekproef van (andere) 20 items als post-test.

In de spreadsheet vanuit Emergo worden de resultaten in de studie WEL in onderstaande volgorde en nummering, en met aanduiding van mini-game aangeboden.

#### Vraag 1 (MG 1)

Waar kunnen we in een organisatie het antwoord vinden op de vragen: 'Waarom bestaan we; waarom doen we wat we doen?'

- A. Doelstellingen
- B. Ambitie
- C. Visie
- D. **Missie**

#### Vraag 2 (MG 1)

Wat moet een organisatie rekening mee houden als het gaat om de missie en visie

- A. **Missie, visie en ambities mogen niet alleen statements zijn in brochures en op websites; ze moeten ook waarneembaar zijn in de organisatie.**
- B. Missie, visie en ambities zijn alleen statements in beleidsdocumenten en brochures
- C. Missie, visie en ambities zijn altijd zichtbaar in de organisatie
- D. Missie, visie en ambities wijken soms af van de werkelijkheid. Dat hoeft niet perse in een lijn te liggen.

#### Vraag 3 (MG 1)

Waar kunnen we in een organisatie het antwoord vinden op de vragen: 'Hoe ziet de toekomstige omgeving van onze organisatie eruit en wat is dan de ideale positie voor onze organisatie?'

- A. **Visie**
- B. Ambitie
- C. Strategie
- D. Missie

#### Vraag 4 (MG 1)

Wat zijn kernwaarden?

- A. De leidende doelstellingen van de organisatie vertaald in aantallen of bedragen.
- B. **De leidende principes, overtuigingen en drijfveren van de organisatie en haar leden.**
- C. De leidende principes die de organisatie door middel van haar missie wil uitdragen.
- D. De leidende principes die de organisatie op middellange termijn ambieert.

#### Vraag 5 (MG 2)

Wat is het doel van een analyse van de DESTEP factoren?

- A. Een beeld krijgen van de interne ontwikkelingen
- B. Een beeld krijgen van de marktaantrekkelijkheid
- C. **Een beeld krijgen van de externe invloeden en ontwikkelingen**
- D. Een beeld krijgen van de marktomgeving

#### Vraag 6 (MG 2)

Eenzaamheid komt vaak voor onder ouderen. Bij welke DESTEP factor past deze ontwikkeling?

- A. **Sociaal-maatschappelijke ontwikkeling**
- B. Demografische ontwikkeling

- C. Politiek-juridische ontwikkeling
- D. Technologische ontwikkelin

*Vraag 7 (MG 2)*

Welke bronnen heb je nodig om een DESTEP analyse te maken?

- A. Documenten die de toekomst van de zorg omschrijven
- B. Beleidsdocumenten
- C. Bewoners van het verzorgingstehuis
- D. Alle bovengenoemde documenten**

*Vraag 8 (MG 2)*

De D van de DESTEP methode staat voor?

- A. Demografische gegevens
- B. Demografische ontwikkelingen**
- C. Duurzame ontwikkelingen
- D. Duurzame prestatie

*Vraag 9 (MG 2)*

Het is soms lastig om een indicatie te krijgen in de zorg. Dat een indicatie nodig is een externe invloed die beschreven kan worden in een DESTEP analyse. Bij welke DESTEP factor past deze ontwikkeling?

- A. Demografische ontwikkeling
- B. Sociaal-maatschappelijke ontwikkeling
- C. Politiek-juridische ontwikkeling**
- D. Technologische ontwikkeling

*Vraag 10 (MG 3)*

Wat analyseer je met het 5 krachten model van Porter?

- A. De marktaantrekkelijkheid van een branche**
- B. De maatschappelijke ontwikkelingen binnen een branche
- C. De kansen en bedreigingen in de omgeving
- D. Hoe de omgeving de organisatie beïnvloed

*Vraag 11 (MG 3)*

Er kunnen verschillende ontwikkelingen zijn in de zorgsector. Een ervan is dat er steeds meer ouderen komen en de markt groeit. Welke kracht van Porter wordt door deze omschrijving beschreven?

- A. Dreiging van substituten
- B. Rivaliteit tussen bestaande aanbieders in de markt**
- C. Onderhandelingsmacht van leveranciers
- D. Dreiging nieuwe toetreders op de markt

*Vraag 12 (MG 3)*

Wat is GEEN kracht binnen het 5krachtenmodel van Porter?

- A. Onderhandelingskracht leveranciers
- B. Rivaliteit tussen bestaande aanbieders
- C. Dreiging nieuwe toetreders op de markt
- D. Dreiging van klanten**

*Vraag 13 (MG 3)*

Er kunnen verschillende ontwikkelingen zijn in de zorgsector. Een ervan is dat ouderen steeds langer thuis wonen en er verschillende technieken zijn om ze te helpen. Welke kracht van Porter wordt door deze omschrijving beschreven?

- A. Dreiging van substituten**
- B. Onderhandelingsmacht van leveranciers
- C. Dreiging nieuwe toetreders op de markt
- D. Onderhandelingsmacht van de afnemers

*Vraag 14 (MG 4)*

Een manager heeft verschillende taken. Het loopt niet zo lekker tussen twee afdelingen binnen een zorginstelling. Een groep medewerkers heeft het gevoel dat ze niet meer gewaardeerd worden. De manager wil dit verbeteren door in gesprek te gaan met deze groep mensen. Welke functie van communicatie gebruikt de manager in dit voorbeeld?

- A. Verbeteren
- B. Informeren
- C. Betrekken
- D. Binden**



*Vraag 15 (MG 4)*

Een voorbeeld van een managementtechniek die een manager kan toepassen is het 'management by objectives'. In welke van de beschreven situaties is er sprake van 'management by objectives'?

- A. De manager vertelt een doel (het invoeren van een vragenlijst), maar laat de uitvoering over aan de verpleegkundigen
- B. De manager overlegt met de verpleegkundige en zij stellen samen een plan vast om de vragenlijst in te voeren. Dit plan wordt daarna gezamenlijk geëvalueerd**
- C. De manager wil graag een vragenlijst invoeren en spreekt met verpleegkundigen over de voor en nadelen van de vragenlijst en vraagt input. Er wordt een positieve werkomgeving gecreëerd.
- D. De manager geeft de taak (invoeren vragenlijst) aan een verpleegkundige. Daarna is laat hij de uitvoering en verantwoording over aan de verpleegkundige.

*Vraag 16 (MG 4)*

Een manager wil een extra bericht plaatsen op intranet om input te vragen aan de verpleegkundige rondom een probleem. Welke functie van communicatie gebruikt de manager in dit voorbeeld?

- A. Verbeteren
- B. Informeren
- C. Betrekken**
- D. Binden

*Vraag 17 (MG 4)*

De locatiemanager geeft in een gesprek aan wat het doel is en laat de invulling hiervan aan de verpleegkundigen over. Welke management techniek wordt hier gebruikt?

- A. management by exception**
- B. management by objectives
- C. management by break-through
- D. management by motivation

*Vraag 18 (MG 5)*

Binnen een organisatie is het handig om te weten wie naar wie luistert. Op deze manier kan je via bepaalde maatregelen macht uitoefenen en anderen beïnvloeden. Welke maatregel is een voorbeeld van het uitoefenen van de deskundigheidsmacht als we het hebben over het invoeren van een vragenlijst?

- A. Iemand vaker inhuren indien deze persoon het invullen van de vragenlijst goed doet.
- B. Het toezeggen om het resultaat van de invoering van de vragenlijst mee te nemen in de beoordeling
- C. Iemand met expertise laten uitleggen waarom het een goed idee is om iets in te voeren.**
- D. Iemand een officiële waarschuwing geven als de vragenlijst niet ingevoerd wordt.

*Vraag 19 (MG 5)*

Hoe wordt een machtsbron die ontleend wordt aan de persoonlijke eigenschappen van een persoon ook genoemd?

- A. Deskundigheidsmacht
- B. Beloningsmacht
- C. Legitieme macht
- D. Charismatische macht**

*Vraag 20 (MG 5)*

Geef het best passende antwoord op de volgende stelling. Het gebruik van een machtsbron in organisaties is altijd toegestaan.

- A. Eens, er zijn nu eenmaal verschillen tussen mensen, dus ook verschillen in macht. Je hebt er hard voor gewerkt om deze machtsbron te kunnen gebruiken. Iedereen heeft wel een vorm van macht.
- B. Eens, maar de maatregelen die genomen worden moeten wel in proportie staan van de taak waarvoor de machtsbron gebruikt wordt.**
- C. Oneens, een machtsbron gebruiken is altijd fout
- D. Oneens, een machtsbron wordt door de hogergeplaatste medewerkers gebruikt, het is oneerlijk om dit actief te gebruiken.

*Vraag 21 (MG 5, post)*

Hoe wordt een machtsbron die gebaseerd is op kennis en argumentatie ook genoemd?

- A. Deskundigheidsmacht
- B. Informatiemacht**
- C. Legitieme macht
- D. Charismatische macht

*Vraag 22 (MG 5, post)*

Binnen een organisatie is het handig om te weten wie naar wie luistert. Op deze manier kan je via bepaalde maatregelen macht uitoefenen en anderen beïnvloeden. Welke maatregel is een voorbeeld van het uitoefenen van de informatiemacht als we het hebben over het invoeren van een vragenlijst?

- A. Iemand toegang geven besluitvormingsdocumenten en communicatie.**
- B. Het toezeggen om het resultaat van de invoering van de vragenlijst mee te nemen in de beoordeling
- C. Iemand met expertise laten uitleggen waarom het een goed idee is om iets in te voeren.
- D. Iemand een officiële waarschuwing geven als de vragenlijst niet ingevoerd wordt.

*Vraag 23 (MG 6)*

In een functieprofiel van een verpleegkundige staat veel informatie. Een voorbeeld hiervan is dat een verpleegkundige verpleegtechnische handelingen uitvoert. Dit is te typeren als:

- A. Taak**
- B. Competentie
- C. Kennis
- D. Bevoegdheid

*Vraag 24 (MG 6)*

Het begrip competentie bevat de volgende dimensies:

- A. Kennis, vaardigheden, ervaring
- B. Kennis, ervaring, houding
- C. Kennis, vaardigheden, houding**
- D. Vaardigheden, houding, ervaring

*Vraag 25 (MG 6, post)*

Bij competentie management is periodiek ruimte ingeruimd om de essentiële competenties en de ontwikkelingsbehoefte van medewerkers te beoordelen. Welke van de volgende instrumenten wordt hiervoor gebruikt.

- A. Assessment**
- B. Training
- C. Coaching
- D. Functieprofiel

*Vraag 26 (MG 6, post)*

Bij competentie management is het zaak om talenten en competenties van medewerkers af te stemmen op de strategische doelen van een organisatie. Daarvoor zijn een drietal soorten activiteiten noodzakelijk. Welke activiteit hoort hier niet bij?

- A. Organisatiebehoefte – de organisatie formuleert haar doelen en noodzakelijke competenties.
- B. Matchen – het ontwikkelen van gerichte ontwikkelingsprogramma's.
- C. Gesprekscyclus – het inrichten en uitvoeren van taakgesprek, beoordelingsgesprek en functioneringsgesprek.**
- D. Individuele wens – de persoonlijke ambities en interesses van medewerkers en op haalbaarheid beoordeeld worden in een potentieel schatting.

*Vraag 27 (MG 7)*

Bij competentie management is de organisatie primair gericht op het belang van medewerkers en managers en dat ze zo optimaal mogelijk functioneren. In de zorginstelling de Kastanjehoeve wordt gevraagd om de competenties van een verpleegkundige te benoemen die nodig zijn bij een vacaturetekst. Er zijn verschillende bronnen waaruit je deze informatie kan halen. Welke bronnen zijn dit?

- A. Oude vacaturetekst
- B. Beleidsdocument
- C. Functieprofiel verpleegkundige
- D. Alle drie de bovenstaande bronnen**

*Vraag 28 (MG 7)*

Een organisatie-indeling kan op verschillende manieren grafisch worden weergegeven in een organigram. De keuze voor een vorm is afhankelijk van het doel van het organigram en de keuze van de organisatie om ergens de nadruk op te leggen. Geef aan welk antwoord niet juist is.

- A. Een verticaal organigram geeft duidelijk aan wie de baas is en hoe de gezagsverhoudingen liggen.
- B. Een horizontaal organigram legt de nadruk niet zo dominant op hiërarchie.

- C. Een concentrisch organigram geeft in centrische cirkels weer hoe van binnen naar buiten de volgorde van hiërarchie is.
- D. **Een vierkant organigram geeft aan hoe in een organisatie die volgens een matrixvorm is ingericht de projecten en functies zich tot elkaar verhouden.**

*Vraag 29 (MG 7)*

Spanwijdte (span of control), spandiepte (depth of control) en omspanningsvermogen (scope of control) zijn begrippen die iets aangeven over de organisatiestructuur. Welke bewerking is *niet* juist?

- A. **Bij spandiepte gaat het over het aantal medewerkers waar effectief leiding aan kan worden gegeven.**
- B. In de meest ideale situatie zijn zowel spandiepte en spanwijdte gelijk aan het omspanningsvermogen.
- C. Bij spanwijdte kan gerefereerd worden aan het aantal afdelingen of het aantal medewerkers.
- D. Een grote spandiepte is vaak aanwezig bij grote ondernemingen, daarbij spreken we ook wel van steile organisaties.

*Vraag 30 (MG 7)*

Een organisatiestructuur kan vastgesteld worden op basis van de verdeling en toepasbaar op mensen in de organisatie. Op basis van welke kenmerken kan de organisatiestructuur beschreven worden?

- A. Taken
- B. Bevoegdheden
- C. **Beslissingen**
- D. Verantwoordelijkheden

*Vraag 31 (MG 7)*

Wat is een spanwijdte (span of control)?

- A. Een spanwijdte is gelijk aan het aantal hiërarchische niveaus waar effectief leiding aan kan worden gegeven.
- B. Een spanwijdte is gelijk aan het aantal afdelingen binnen de organisatie.
- C. **Een spanwijdte is gelijk aan aantal medewerkers waaraan een manager direct leiding geeft.**
- D. Een spanwijdte is gelijk aan het aantal medewerkers waaraan een manager effectief leiding kan geven.

*Vraag 32 (MG 8)*

Bij het maken van een stroomschema maak je gebruik van verschillende symbolen om het totale proces te visualiseren. Een van die symbolen is een cirkel. Wat betekent dit?

- A. **Start of eindpunt van het proces**
- B. Een beslissingsmoment
- C. Een activiteit
- D. Een formulier of document

*Vraag 33 (MG 8)*

Naast een organisatiestructuur is er ook een structuur rondom (kern)activiteiten binnen de organisatie (een processtructuur) te onderscheiden. Geef aan waar de processtructuur voornamelijk op is gericht. Welk antwoord is niet juist?

- A. Het leveren van een resultaat.
- B. Het leveren van de output van een proces.
- C. Het maken van afspraken over uitvoering en besturing.
- D. **Activiteiten en taken, verantwoordelijkheden en bevoegdheden beschrijven.**

*Vraag 34 (MG 8)*

Er is een probleem met de maaltijdistributie. Het is nodig om het huidige proces van de maaltijdistributie te bekijken. Hoe pak je dit aan?

- A. Zoek een oud stroomschema op waarin het maaltijdistributie proces is beschreven.
- B. Maak een stroomschema door met de verschillende medewerkers te praten.
- C. Bekijk verschillende beleidsdocumenten.
- D. **Alle drie de stappen zijn nodig.**

*Vraag 35 (MG 8, post)*

Bij het maken van een stroomschema maak je gebruik van verschillende symbolen om het totale proces te visualiseren. Een van die symbolen is een ruitvorm. Wat betekent dit?

- A. Start of eindpunt van het proces
- B. Een beslissingsmoment**
- C. Een activiteit
- D. Een formulier of document

*Vraag 36 (MG 8)*

Bij het inrichten van processen is het belangrijk na te denken over een logische volgorde van activiteiten en toewijzing van betrokkenen bij de diverse activiteiten. Hiervoor zijn diverse schematechnieken die daarvoor gebruikt kunnen worden. In een zorginstelling kan bijvoorbeeld de maaltijdistributie niet optimaal werken. Met welk soort schema is deze maaltijdistributie inzichtelijk te maken, met mogelijkheden tot analyse en verbetering?

- a. Gantt-chart oftewel strokenplanning
- b. Stroomschema**
- c. Netwerkplanningsschema
- d. Processtructuur overzichtsschema

*Vraag 37 (MG 9)*

Wat betekenen de letters van de een RASCI-matrix

- A. R (responsible), A (Accountable), S (Supportive), C (Consulted), I (informed)**
- B. R (responsible), A (Accepted), S (Supportive), C (Controlled), I (informed)
- C. R (responsible), A (Accepted), S (Supportive), C (Consulted), I (informed)
- D. R (responsible), A (Accountable), S (Supportive), C (Controlled), I (informed)

*Vraag 38 (MG 9)*

Als er iets misgaat in de proces, is het goed om het probleem eerst te achterhalen door middel van de procesrollen inzichtelijk te maken. Dit kan je op verschillende manieren doen. Welke analyse is hierbij toepasbaar?

- A. Macht analyse
- B. RASCI matrix**
- C. DESTEP
- D. Porter analyse

*Vraag 39 (MG 9)*

In een proces hebben medewerkers verschillende rollen. Bij een van die rollen moet de medewerker vooraf advies geven over een activiteit. Welke benaming volgens het RASCI model past hierbij?

- A. Informed
- B. Consulted**
- C. Supportive
- D. Responsible

*Vraag 40 (MG 9)*

Binnen een zorginstelling hebben verschillende soorten medewerkers verschillende rollen per processtap. Dit kan worden gevisualiseerd in een zogeheten RASCI-matrix. Deze begrippen hebben elk een specifieke procesrol. Welk antwoord is een correcte weergave deze rollen?

- A. R (Responsible) > rol: proceseigenaar; S (Supportive) > rol: accepteren.
- B. A (Accountable) > rol: beoordelen; I (Informed) > rol: informeren.
- C. A (Accountable) > rol: aanspreken; S (Supportive) > rol: archiveren.**
- D. R (Responsible) > rol: proceseigenaar; C (Consulted) > rol: accepteren.



## Appendix 1.2. Pre/post test vragen Junior Scientist

### Vraag 1 (MG 0: worry questions)

Stel je krijgt onderstaand fragment over een bepaald onderzoek te lezen:

‘Er is vragenlijstonderzoek gedaan naar het verband tussen digitalemediagebruik en het mentale welzijn van ouderen. Het verband tussen digitalemediagebruik en mentaal welzijn toont aan dat digitalemediagebruik invloed heeft op het mentale welzijn van ouderen’

Welke stelling is op deze conclusie het meest van toepassing?

- A. **Een vragenlijstonderzoek kan wel een verband laten zien, maar niet de conclusie van een causaal verband ondersteunen**
- B. Zonder de steekproefgrootte te kennen kan wel een verband gevonden worden, maar geen causaal verband
- C. Het is onwaarschijnlijk dat er een verband is tussen digitalemediagebruik en mentaal welzijn
- D. Een vragenlijstonderzoek is niet geschikt om een verband tussen digitalemediagebruik te onderzoeken

### Vraag 2 (MG 0)

Stel je krijgt onderstaand fragment over een bepaald onderzoek te lezen:

‘Een peiling onder de bevolking wees uit dat slechts 30% van de ondervraagden positief was over de politieke kandidaat’

Waarom is deze samenvatting misleidend?

- A. **Er wordt niet duidelijk gemaakt wat precies de antwoordopties waren, waardoor niet duidelijk is hoeveel ondervraagden negatief waren over de politieke kandidaat**
- B. De redenen om positief over de kandidaat te zijn ontbreken, waardoor niet duidelijk is in hoeverre positief ook echt positief is
- C. De samenvatting mist een statistische toets
- D. Doordat alleen ondervraagden zijn opgenomen in het onderzoek is er sprake van een validiteitsbedreiger, namelijk selectie.

### Vraag 3 (MG 0)

Stel je krijgt onderstaand fragment over een bepaald onderzoek te lezen:

‘Er is vragenlijstonderzoek gedaan naar het verband tussen de omvang van een sociaal netwerk en de mate van levensgeluk onder adolescenten. Het verband tussen het sociaal netwerk en levensgeluk toont aan dat de omvang van het sociaal netwerk invloed heeft op het levensgeluk van adolescenten’

Welke stelling is op deze conclusie het meest van toepassing?

- A. **Een vragenlijstonderzoek kan wel een verband laten zien, maar niet de conclusie van een causaal verband ondersteunen**
- B. Zonder de steekproefgrootte te kennen kan wel een verband gevonden worden, maar geen causaal verband
- C. Het is onwaarschijnlijk dat er een verband is tussen het sociaal netwerk en levensgeluk
- D. Een vragenlijstonderzoek is niet geschikt om een verband tussen het sociaal netwerk en levensgeluk te onderzoeken

### Vraag 4 (MG 0)

Stel je krijgt onderstaand fragment over een bepaald onderzoek te lezen:

‘Een peiling onder de bevolking wees uit dat slechts 25% van de ondervraagden positief waren over de aangekondigde wetswijziging’

Waarom is deze samenvatting misleidend?

- A. **Er wordt niet duidelijk gemaakt wat precies de antwoordopties waren, waardoor niet duidelijk is hoeveel ondervraagden negatief waren over de aangekondigde wetswijziging**
- B. De redenen om positief over de aangekondigde wetswijziging te zijn ontbreken, waardoor niet duidelijk is in hoeverre positief ook echt positief is
- C. De samenvatting mist een statistische toets

- D. Doordat alleen ondervraagden zijn opgenomen in het onderzoek is er sprake van een validiteitsbedreiger, namelijk selectie.

*Vraag 5 (MG 1: onderzoeksdesign)*

Een onderzoeker wil de Kwaliteit van Leven van ouderen in kaart brengen. Zij benadert deze via de ouderencentra. Welke bedreiger van externe validiteit speelt hier een rol?

- A. **Niet-representatieve steekproeven**
- B. Groei
- C. Instrumentatie
- D. Reactieve experimentele locatie

*Vraag 6 (MG 1)*

Stelling 1: 'Als je meer variabelen meeneemt en toetst in je onderzoek, dan heb je ook een grotere steekproef nodig.

Stelling 2: 'Gegeven een beperkte steekproefomvang moet je daarom selectief zijn in welke controlevariabelen je mee gaat nemen.'

Welke stelling is juist?

- A. **Stelling 1 is juist en Stelling 2 is juist**
- B. Stelling 1 is onjuist en Stelling 2 is onjuist
- C. Stelling 1 is juist, stelling 2 is onjuist
- D. Stelling 2 is onjuist, stelling 2 is juist

*Vraag 7 (MG 1)*

Een onderzoeker wil de attitude jegens koffie onder adolescenten in kaart brengen. Zij benadert deze via koffietentjes.

Welke bedreiger van externe validiteit speelt hier een rol?

- A. **Selectie**
- B. Groei
- C. Testeffect
- D. Reactieve experimentele locatie

*Vraag 8 (MG 1)*

Stelling 1: 'Als je meer controlevariabelen opneemt in je onderzoek mag je een kleinere steekproef hebben.'

Stelling 2: 'Gegeven een beperkte steekproefomvang moet je daarom selectief zijn in welke controlevariabelen je mee gaat nemen.'

Welke stelling is juist?

- A. **Stelling 1 is onjuist en Stelling 2 is juist**
- B. Stelling 1 is juist en Stelling 2 is onjuist
- C. Stelling 1 is juist, stelling 2 is onjuist
- D. Stelling 2 is onjuist, stelling 2 is juist

*Vraag 9 (MG 2: data verzameling)*

Wat is het belangrijkste voordeel van het bijhouden van een bijzonderhedendocument of labjournaal tijdens een dataverzameling?

- A. Het stelt andere onderzoekers in staat om het databestand zelfstandig te gebruiken
- B. **Het helpt om afwijkingen in een databestand te herkennen die tijdens de dataverzameling zijn ontstaan en er gericht besluiten over te kunnen nemen**
- C. Het bijhouden van een bijzonderhedendocument of labjournaal is een verouderd gebruik en heeft tegenwoordig geen voordelen meer
- D. Het wordt gebruikt om proefpersonen een op-maat-gemaakte debriefing te geven

*Vraag 10 (MG 2)*

Wat is het belangrijkste voordeel van het bijhouden van een bijzonderhedendocument of labjournaal tijdens een dataverzameling?

- A. Het geldt als bewijs dat de data ook echt door onderzoek verkregen zijn
- B. **Het stelt de onderzoeker in staat om problemen met de data beter te begrijpen en op te lossen**
- C. Het bijhouden van een bijzonderhedendocument of labjournaal is een verouderd gebruik en heeft tegenwoordig geen voordelen meer
- D. Het wordt gebruikt om proefpersonen een op-maat-gemaakte debriefing te geven

Vraag 11 (MG 3: data analyse)

Uitbijters zijn...

- A. alle observaties die statistische ver afwijken van het centrum van een dataverdeling
- B. observaties in de steekproef die niet tot de beoogde populatie van het onderzoek behoren**
- C. invoerfouten
- D. verwaarloosbaar zolang  $N > 30$

Vraag 12 (MG 3)

Welke uitspraak is juist?

- A. Uitbijters kunnen vrijwel altijd het beste verwijderd worden uit een dataset
- B. Er moet een goede reden zijn om uitbijters te verwijderen uit een dataset**
- C. Uitbijters mogen nooit uit een dataset verwijderd worden
- D. Het verwijderen van uitbijters moet worden bijgehouden in een labjournaal

Vraag 13 (MG 3)

Welke uitspraak is juist?

- A. Bij een  $N > 30$  hebben invoerfouten nauwelijks invloed op de resultaten
- B. Invoerfouten kunnen de resultaten soms behoorlijk vertekenen**
- C. Nu data via computers verkregen worden zijn invoerfouten onmogelijk
- D. Een invoerfout mag niet aangepast worden

Vraag 14 (MG 3)

Wat is een assumptie van een lineair model?

- A. Normale verdeling van de residuen**
- B. Normale verdeling van de afhankelijke variabele
- C. Onafhankelijkheid van de afhankelijke variabele
- D. Dat residuen gecorreleerd zijn

Vraag 15 (MG 3)

Welke situatie schendt een assumptie van covariaten:

- A. De score op de covariaat verschilt significant tussen experimentele condities**
- B. De covariaat is negatief geassocieerd met de afhankelijke variabele
- C. Residuen zijn niet gecorreleerd
- D. Er is géén interactieeffect tussen manipulatie en covariaat op de afhankelijke variabele

Vraag 16 (MG 3)

Wat is de naam van een design met zowel tussen- als binnenproefpersonenfactoren?

- A. Een mixed design ANOVA**
- B. Een factoriele ANOVA
- C. Een t-toets
- D. Een pre-experiment

Vraag 17 (MG 3)

Om uitbijters te detecteren kun je een z-waarde van  $z > |3.29|$  hanteren of naar een boxplot kijken. Boxplots hanteren vaak 1.5 keer de interkwartielafstand ( $1.5 \cdot IQR = \text{interquartile range}$ ) als cut-off.

Wat is de strengere cut-off bij een perfecte normaalverdeling van de data?

- A. Een z-waarde  $> |3.29|$  hanteert een strengere cut-off voor uitbijters dan boxplots ( $1.5 \cdot IQR$ )**
- B. Boxplots ( $1.5 \cdot IQR$ ) hanteren een strengere cut-off voor uitbijters dan een z-waarde  $> |3.29|$
- C. Beide cutoffs zijn even streng
- D.  $1.5 \cdot IQR$  bestaat niet niet als uitbijtercriterium

*Vraag 18 (MG 3)*

Stel je onderzoeksvraag is:

'Vergroot 15 minuten mindfulness-oefeningen werknemerbevlogenheid meer dan 15 minuten oefeningloze pauze, en is dit effect sterker voor vrouwen dan voor mannen?'

Hoe kun je deze vraagstelling het beste toetsen?

- A. **Factoriele anova met conditie en geslacht als fixed factors**
- B. t-toetsen voor conditie en geslacht, en een factoriele anova om de interactie conditie\*geslacht te toetsen
- C. Vragenlijstonderzoek
- D. Een t-toets met conditie en werknemerbevlogenheid voor de groep mannen, en een t-toets voor de groep vrouwen.

*Vraag 19 (MG 3)*

In een ancova moet eerst getoetst worden of er een interactie is tussen covariaat en manipulatie.

Welke assumptie van ancova's beschrijft dit?

- A. De assumptie van homogene regressie-hellingshoeken (regression slopes)
- B. De assumptie van lineariteit van het verband
- C. De assumptie van onafhankelijk van covariaat en manipulatie
- D. De assumptie van ongecorreleerde residuen

*Vraag 20 (MG 3)*

Een verschil tussen groepen met een p-waarde van  $p = .003$  betreft een sterker effect dan een verschil met een p-waarde van  $p = .04$ .

Is deze uitspraak juist of onjuist?

- A. juist, want des te kleiner de p-waarde des te sterker een effect
- B. **onjuist, want de omvang van de p-waarde informeert niet over de effectgrootte**
- C. juist, want het verschil is significant
- D. onjuist, want  $.003$  is kleiner dan  $.04$ , dus het effect is juist kleiner

*Vraag 21 (MG 3)*

Stelling 1: Uitbijters zijn alle observaties die statistische ver afwijken van het centrum van een dataverdeling

Stelling 2: Uitbijters zijn observaties in de steekproef die niet tot de beoogde populatie van het onderzoek behoren

Welke stelling is juist?

- A. Stelling 1 is juist, en stelling 2 is juist
- B. **Stelling 1 is onjuist, en Stelling 2 is juist**
- C. Stelling 1 is juist, en stelling 2 is onjuist
- D. Stelling 1 is onjuist, en stelling 2 is onjuist

*Vraag 22 (MG 3)*

Een onderzoeker heeft met een statistische methode een aantal extreme waarden geïdentificeerd. Vervolgens heeft de onderzoeker alle extreme waarden zondermeer verwijderd. Is dit een goed idee geweest?

- A. Ja, want uitbijters kunnen vrijwel altijd zondermeer verwijderd worden uit een dataset
- B. **Nee, er moet een goede reden zijn om uitbijters te verwijderen uit een dataset**
- C. Nee, want uitbijters mogen nooit uit een dataset verwijderd worden
- D. Nee, want verwijderen van uitbijters moet worden bijgehouden in een labjournaal

*Vraag 23 (MG 3)*

Welke uitspraak is juist?

- A. Bij een  $N > 30$  hoeft een onderzoeker invoerfouten niet aan te passen of te verwijderen
- B. **Voor het aanpassen of verwijderen van invoerfouten moet je proberen om de meest ruwe data raad te plegen**
- C. Nu data via computers verkregen worden zijn invoerfouten onmogelijk
- D. Een invoerfout mag niet aangepast worden

*Vraag 24 (MG 3)*

Wat is een assumptie van een lineair model?

- A. **Een lineair verband tussen predictoren en afhankelijke variabele**
- B. Normale verdeling van de afhankelijke variabele
- C. Dat residuen niet uitmiddelen tot nul

- D. Dat residuen gecorreleerd zijn

*Vraag 25 (MG 3)*

Welke situatie schendt een assumptie van covariaten:

- A. **Het verband tussen covariaat en afhankelijke variabele is afhankelijk van de conditie in de experimentele manipulatie**
- B. De covariaat is geassocieerd met de afhankelijke variabele
- C. Residuen zijn gecorreleerd
- D. De covariaat is negatief geassocieerd met de afhankelijke variabele

*Vraag 26 (MG 3)*

In een bedrijf worden teams toegewezen aan een controle of experimentele conditie. De onderzoeker wilt vervolgens weten of de manipulatie invloed had op de arbeidsatisfactie van individuele teamleden? Dit is een voorbeeld van...

- A. **Een quasi-experiment**
- B. Een gerandomiseerd experiment
- C. Een mixed-design ANOVA
- D. Een descriptief onderzoek

*Vraag 27 (MG 3)*

Het is mogelijk om uitbijters te detecteren door scores naar z-scores om te zetten, en alle z-waarden extremer dan -3.29 of 3.29 als extreem te bestempelen.

Het is ook mogelijk om uitbijters te detecteren door de interkwartielafstand rondom de mediaan te bepalen. Waarden extremer dan 1.5x de interkwartielafstand kunnen als extreem bestempeld worden.

Welke methode heeft de voorkeur als de variabele erg scheef verdeeld is?

- A. **De interkwartielafstand**
- B. De z-score
- C. Bij een scheve verdeling is zoeken naar uitbijters niet meer mogelijk
- D. Visuele inspectie

*Vraag 28 (MG 3)*

Stel je onderzoeksvraag is:

‘Leidt het online volgen van hoorcolleges tot minder studiemotivatie dan het fysiek bijwonen van hoorcolleges, en is dit effect sterker voor vrouwen dan voor mannen?’

Hoe kun je deze vraagstelling het beste toetsen?

- A. **Factoriele anova met conditie en geslacht als fixed factors**
- B. t-toetsen voor conditie en geslacht, en een factoriele anova om de interactie conditie\*geslacht te toetsen
- C. Vragenlijstonderzoek
- D. Een t-toets met conditie en studiemotivatie voor de groep mannen, en een t-toets voor de groep vrouwen

*Vraag 29 (MG 3)*

In een ancova moet eerst getoetst worden of de scores op de covariaat gemiddeld verschillen tussen de experimentele condities.

Welke assumptie van ancova's beschrijft dit?

- A. **De assumptie dat een covariaat geen verband heeft met de manipulatie**
- B. De assumptie van homogene regressie-hellingshoeken (regression slopes)
- C. De assumptie van lineariteit van het verband
- D. De assumptie van ongecorreleerde residuen

*Vraag 30 (MG 3)*

Een onderzoeker vindt een p-waarde van  $p = .001$  en een p-waarde van  $.04$ . In het verslag schrijft de onderzoeker:

“Het eerste verschil was zeer significant ( $p = .001$ ), en het tweede verschil was maar een beetje significant ( $p = .04$ ).”

Is dit een correcte interpretatie van p-waarden?

- A. Ja, want des te kleiner de p-waarde des te sterker een effect
- B. Nee, want de omvang van de p-waarde informeert niet over de 'kracht' van significantie**
- C. Ja, want het verschil is significant
- D. Nee, want .001 is een beetje significant, en .04 zeer significant

*Vraag 31 (MG 4: rapportage)*

De interventiegroep scoorde hoger ( $M = 4.93$ ;  $SD = 1.01$ ) op prestatiemotivatie dan de controlegroep ( $M = 3.91$ ;  $SD = 1.03$ ),  $t(422) = -10.36$ ,  $p < .001$ .

Wat ontbreekt in deze rapportage om het effect te kunnen duiden?

- A. de effectgrootte**
- B. het aantal deelnemers
- C. De naam van de toets die is uitgevoerd
- D. De richting van het verband

*Vraag 32 (MG 4)*

Welke uitspraak is juist?

- A. Een figuur moet je altijd in de hoofdtekst toegelicht worden**
- B. Figuren moeten op zichzelf kunnen staan, en hoeven daarom niet in de hoofdtekst toegelicht te worden
- C. Een figuur zou nooit nodig moeten zijn
- D. Een figuur vertelt meer dan duizend woorden, dus een verslag kan nooit genoeg figuren hebben

*Vraag 33 (MG 4: rapportage)*

De interventiegroep scoorde hoger ( $M = 4.93$ ;  $SD = 1.01$ ) op prestatiemotivatie dan de controlegroep ( $M = 3.91$ ;  $SD = 1.03$ ),  $F(1,422) = -20.87$ ,  $p < .001$ .

Wat ontbreekt in deze rapportage om het effect te kunnen duiden?

- A. de effectgrootte**
- B. het aantal deelnemers
- C. De naam van de toets die is uitgevoerd
- D. De richting van het verband

*Vraag 34 (MG 4)*

Welke uitspraak is juist?

- A. Een figuurbijschrift moet vooral beknopt zijn
- B. Figuren moeten op zichzelf kunnen staan, dus het bijschrift de figuur beknopt doch volledig begrijpelijk maken**
- C. Een figuur zou nooit nodig moeten zijn
- D. Een figuur vertelt meer dan duizend woorden, dus een figuur toevoegen overschrijdt al snel de woordlimiet

*Vraag 35 (MG 5: discussie)*

In een gerandomiseerd onderzoeksontwerp is een voormeting noodzakelijk.

Deze stelling is:

- A. Niet waar, want randomisatie zou systematische voormetingverschillen moeten voorkomen**
- B. Waar, want zonder voormeting kan het geen experiment zijn
- C. Niet waar, want bij randomisatie kan geen voormeting worden toegepast
- D. Waar, want een voormeting wordt gebruikt om te randomiseren

*Vraag 36 (MG 5)*

Stel je leest onderstaand fragment in een discussiesectie van een wetenschappelijk artikel over het welzijn onder ouderen: 'Een tekortkoming van dit onderzoek is dat de vragenlijst onder ouderen op een computer is afgenomen. Daardoor zijn ouderen zonder computerkennis niet opgenomen in de steekproef.'

Heeft de auteur hier voldoende toelichting gegeven?

- A. Nee, de auteur moet ook toelichten waarom computerkennis de conclusies kan nuanceren**
- B. Ja, de auteur hoeft alleen maar suggesties te geven voor waar toekomstig onderzoek zich op zou kunnen richten
- C. Ja, het punt is zelfevident
- D. Nee, de statistische informatie uit de resultaten ontbreken hier

*Vraag 37 (MG 5)*

‘Uit de resultaten van dit gerandomiseerde experiment blijkt dat de bewegingsinterventie effect heeft gehad op de ontwikkeling van executieve functies van kinderen en uit eerder onderzoek blijkt dat executieve functies correleren met intelligentie.’

Kun je dan op basis van deze stellingen concluderen dat de bewegingsinterventie ook een causaal effect heeft op intelligentie?

- A. Ja
- B. Nee**
- C. Ja, maar alleen als de correlatie tussen executieve functies en intelligentie positief is
- D. Moeilijk te zeggen, omdat er niet gerandomiseerd is

*Vraag 38 (MG 5)*

In een niet-gerandomiseerd onderzoeksontwerp is een voormeting van groot belang.

Deze stelling is:

- A. Waar, want zonder randomisatie kunnen er systematische voormetingverschillen zijn**
- B. Waar, want zonder voormeting kan het geen experiment zijn
- C. Niet waar, want zonder randomisatie kan geen voormeting worden toegepast
- D. Waar, want een voormeting is altijd van groot belang

*Vraag 39 (MG 5)*

Stel je leest onderstaand fragment in een discussie-sectie van een wetenschappelijk artikel over het welzijn onder vluchtelingen:

‘Een tekortkoming van dit onderzoek is dat de vragenlijst onder vluchtelingen is afgenomen in aanwezigheid van een tolk. De tolk kan invloed hebben gehad op de antwoorden van de vluchtelingen.’

Heeft de auteur hier voldoende toelichting gegeven?

- A. Nee, de auteur moet ook toelichten waarom de aanwezigheid van de tolk de conclusies kan nuanceren**
- B. Ja, de auteur hoeft alleen maar suggesties te geven voor waar toekomstig onderzoek zich op zou kunnen richten
- C. Ja, het punt is zelfevident
- D. Nee, de statistische informatie uit de resultaten ontbreken hier

*Vraag 40 (MG 5)*

‘Uit de resultaten van dit zuiver experiment blijkt dat de voorlichting over gehoorschade effect heeft gehad op het gebruik van gehoorbescherming bij concerten. Uit eerder onderzoek is gebleken dat het gebruik van gehoorbescherming correleert met een consciëntieuze persoonlijkheidstrek’

Kun je dan op basis van deze stellingen concluderen dat de interventie ook een causaal effect heeft op gebruik van gehoorbescherming via de mate van consciëntieusheid als persoonlijkheidstrek?

- A. Ja
- B. Nee**
- C. Ja, maar alleen als de correlatie tussen gehoorbescherming en consciëntieusheid positief is
- D. Moeilijk te zeggen, omdat er niet gerandomiseerd is

## Appendix 2. (Sub)schalen waarderingsvragenlijst

### Nota Bene

Alle items worden gescoord op een 5-punts Likertschaal met de volgende waarden:

- 1 = zeer mee oneens
- 2 = mee oneens
- 3 = neutraal
- 4 = mee eens
- 5 = zeer mee eens

De selectie van subschalen (vanuit originele instrumenten) is door Rob/Hans gedaan. Hans heeft waar nodig de items vertaald. Voor de Gamebrics studie worden items in het NL aangeboden.

Studenten krijgen de items aangeboden in een willekeurige volgorde, en zien daarbij NIET tot welke (sub)schaal ze horen (ook weer om mogelijke bias bij scoring te voorkomen).

In de spreadsheet vanuit Emergo met de scores komen de resultaten WEL in onderstaande volgorde (met dezelfde nummering), en met aanduiding van de (sub)schaal, te staan. Voor data-analyse zullen ze waar nodig 'omgepooled' worden.

Studenten in de experimentele conditie (met Gamebrics dashboard) krijgen alle 43 items, incl. items over dashboard. Studenten in de controle conditie (zonder Gamebrics dashboard) krijgen slechts 27 items te beantwoorden. De items die specifiek over dashboard gaan zijn als laatste genummerd, het betreft: 6 items schaal 2.1 (28-33); 7 items schaal 2.2 (34-40); en 3 items schaal 2.3.

### Schaal 1: Motivatie

#### Subschaal 1.1: Motivatie – ervaren druk

1. I did not feel nervous at all while doing this task /  
Ik voelde me niet nerveus tijdens deze taak
2. I felt very tense while doing this task /  
Ik voelde me erg gespannen tijdens deze taak
3. I was very relaxed in doing this task /  
Ik was zeer relaxed tijdens deze taak
4. I was anxious while working on this task /  
Ik was zeer ongerust bij het werken aan deze taak
5. I felt pressured while doing this task /  
Ik voelde me onder druk gezet tijdens deze taak

#### Subschaal 1.2: Motivatie- ervaren competentie

6. I think I am pretty good at this task /  
Ik denk dat ik vrij goed presteerde op deze taak
7. I think I did pretty well at this task, compared to other students /  
Ik denk dat ik deze taak vrij goed heb gedaan, in vergelijking met andere studenten
8. After working at this task for a while, I felt pretty competent /  
Na een tijdje aan deze taak werken, voelde ik mij best wel competent
9. I am satisfied with my performance at this task /  
Ik ben tevreden met mijn prestaties op deze taak
10. I was pretty skilled at this task /  
Ik was behoorlijk vaardig tijdens deze taak
11. This was a task I couldn't do very well /  
Dit was een taak die ik niet erg goed kon uitvoeren

### Schaal 2: Feedback

#### Subschaal 2.1: Feedback Acceptatie

12. I think the received feedback at gameplay was correct /  
Ik denk dat de ontvangen feedback tijdens gameplay correct was
13. I do not agree with the feedback received at gameplay /  
Ik ben het niet eens met ontvangen feedback tijdens gameplay
14. Received feedback at gameplay was a correct evaluation of my performance /  
Ontvangen feedback tijdens gameplay was een correcte evaluatie van mijn prestaties



15. It is hard to take the received feedback at gameplay seriously /  
Het is lastig om de tijdens gameplay ontvangen feedback serieus te nemen

16. The feedback at gameplay helped me improve my performance /  
De feedback tijdens gameplay hielp me mijn prestaties te verbeteren

17. I felt quite good with the performance feedback at gameplay /  
Ik voelde me best goed over mijn prestaties tijdens gameplay

28. I think received feedback in the dashboard was correct /  
Ik denk dat de ontvangen feedback in het dashboard correct was

29. I do not agree with the feedback received in the dashboard /  
Ik ben het oneens met de ontvangen feedback in het dashboard

30. Received feedback in the dashboard was a correct evaluation of my performance /  
Ontvangen feedback in het dashboard was een correcte evaluatie van mijn prestaties

31. It is hard to take the received feedback in the dashboard seriously /  
Het is lastig om de ontvangen feedback in het dashboard serieus te nemen

32. The feedback in the dashboard helped me improve my performance /  
De feedback in het dashboard hielp mij de prestaties te verbeteren

33. I felt quite good with the performance feedback in the dashboard /  
Ik voelde me best goed over de feedback op mijn prestaties in het dashboard

#### **Subschaal 2.2: Feedback Begrip**

18. I think I will be using the feedback at gameplay regularly /  
Ik denk dat ik de feedback tijdens gameplay regelmatig ga gebruiken

19. Thought the feedback at gameplay was unnecessarily complex /  
Ik vond de feedback tijdens gameplay onnodig ingewikkeld

20. Thought the feedback at gameplay was easy to use /  
De feedback tijdens gameplay was eenvoudig te gebruiken

21. Think I need support to understand the feedback at gameplay /  
Ik denk dat ik ondersteuning nodig heb om de feedback tijdens gameplay te begrijpen

22. Think the various feedback functions at gameplay are well integrated /  
Ik denk dat verschillende feedback functie tijdens gameplay goed zijn geïntegreerd

23. Think there are many contradictions in the feedback at gameplay /  
Ik denk dat er veel tegenspraak zit in de feedback tijdens gameplay

24. Felt confident to use the feedback at gameplay  
Voelde me vol vertrouwen bij gebruik feedback tijdens gameplay

34. I think I will be using the feedback in the dashboard regularly /  
Ik denk dat ik de feedback in het dashboard regelmatig ga gebruiken

35. Thought the feedback in the dashboard was unnecessarily complex /  
Ik vond de feedback in het dashboard onnodig ingewikkeld

36. Thought the feedback in the dashboard was easy to use /  
De feedback in het dashboard was eenvoudig te gebruiken

37. Think I need support to understand the feedback in the dashboard /  
Ik denk dat ik ondersteuning nodig heb om de feedback in het dashboard te begrijpen

38. Think the various feedback functions in the dashboard are well integrated /  
Ik denk dat verschillende feedback functie in het dashboard goed zijn geïntegreerd

39. Think there are many contradictions in the feedback in the dashboard /  
Ik denk dat er veel tegenspraak zit in de feedback in het dashboard

40. Felt confident to use the feedback in the dashboard /  
Voelde me vol vertrouwen bij gebruik feedback in het dashboard

#### **Subschaal 2.3: Feedback Behoefte**

25. At gameplay there should be more feedback on how to improve yourself /  
Tijdens gameplay moet er meer feedback gegeven over hoe je te verbeteren

26. At gameplay the provided feedback should be less elaborate /

Tijdens gameplay moet de feedback minder uitgebreid zijn

27. At gameplay provided feedback should contain less corrective feedback /

Tijdens gameplay moet de feedback minder correctief zijn

41. In the dashboard there should be more feedback on how to improve yourself /

Tijdens gameplay moet er meer feedback gegeven over hoe je te verbeteren

42. In the dashboard the provided feedback should be less elaborate /

In het dashboard moet de feedback minder uitgebreid zijn

43. In the dashboard the provided feedback should contain less corrective feedback /

In het dashboard moet de feedback minder correctief zijn