

MASTER'S THESIS

**Hoe intuïtief is de notatie van de PGA- modelleertaal?
et opzetten van een experimentele evaluatie**

van der Meer, L. (Leen)

Award date:
2021

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 09. Sep. 2021

Open Universiteit
www.ou.nl



Hoe intuïtief is de notatie van de PGA- modelleertaal?

Het opzetten van een experimentele evaluatie

Opleiding: Open Universiteit, faculteit Management, Science & Technology
Masteropleiding Business Process Management & IT

Programme: Open University of the Netherlands, faculty of Management, Science &
Technology
Master Business Process Management & IT

Cursus: IM9806 Afstudeeropdracht Business Process Management and IT

Student: Leen van der Meer

Identiteitsnummer:

Datum: 23 januari 2021

Afstudeerbegeleider: Ben Roelens

Meelezer: Harry Martin

Versienummer: 20201224

Status: Opgeleverd

Abstract

Dit onderzoek heeft als doel om de intuïtiviteit van twee verschillende versies van de *Process Goal Alignment* (PGA) notatie te evalueren binnen een praktische bedrijfscontext. In het onderzoek van Roelens en Bork (2020) zijn een aantal voorstellen gedaan ter verbetering van de oorspronkelijke PGA-notatie. Het vaststellen of deze verbeteringen aan de intuïtiviteit van de PGA-notatie ook significant zijn voor potentiële eindgebruikers wordt onderzocht in dit rapport.

De onderzoeksvraag is: *“Hoe kan de intuïtiviteit van twee verschillende versies van de PGA-notatie geëvalueerd worden door het uitvoeren van een experiment binnen een praktische bedrijfscontext?”*.

Er is gekozen voor het uitvoeren van een experiment binnen een praktische bedrijfscontext door op basis van de literatuur een experimentele methode op te stellen om de intuïtiviteit tussen twee verschillende versies PGA-modelleertaal te evalueren. Daaruit blijkt dat de oorspronkelijke methodes niet voorzien in het meten van de intuïtiviteit tussen twee verschillende versies van conceptuele modelleertalen en deze te evalueren. Om de intuïtieve begrijpelijkheid te meten is het experiment uitgebreid met het meten van de representatievoorkeur. Uit het onderzoek blijkt dat de visuele wijzigingen aan de nieuwe PGA-elementen onvoldoende zijn om aan te tonen dat deze intuïtiever zijn dan de oorspronkelijke PGA-elementen.

Sleutelbegrippen

Process Goal Alignment methode, intuïtiviteit, conceptueel model, experimentele evaluatie

Samenvatting

Dit onderzoek heeft als doel om de intuïtiviteit van twee verschillende versies van de *Process Goal Alignment* (PGA) notatie te evalueren binnen een praktische bedrijfscontext. Bork et al. (2019) beschrijven een evaluatietechniek om de intuïtiviteit van een modelleertaal te beoordelen. Deze evaluatietechniek is succesvol toegepast op het onderzoek van Roelens en Bork (2020). Dit heeft geleid tot een aantal voorgestelde verbeteringen van de oorspronkelijke PGA-notatie. Het vaststellen of deze verbeteringen aan de intuïtiviteit van de PGA-notatie ook significant zijn voor potentiële eindgebruikers wordt in dit rapport onderzocht.

De probleemstelling is: *“De intuïtiviteit van de nieuwe versie van de PGA-notatie is niet geëvalueerd door het uitvoeren van het experiment binnen een praktische bedrijfscontext”.*

De overkoepelende onderzoeksvraag is: *“Hoe kan de intuïtiviteit van twee verschillende versies van de PGA-notatie geëvalueerd worden door het uitvoeren van een experiment binnen een praktische bedrijfscontext?”.*

Het experiment onderzoekt of de onafhankelijke variabelen (i.e. oorspronkelijke PGA-notatie, nieuwe PGA-notatie) van invloed kunnen zijn op de afhankelijke (i.e. interpretatieve effectiviteit, interpretatieve efficiëntie en Intuïtieve begrijpelijkheid). Om de afhankelijke variabelen te meten zijn de volgende hypothesen opgesteld:

- $H_{\text{effectiviteit}}$: de interpretatieve effectiviteit van de nieuwe PGA-notatie voor element x tijdens taak y is hoger dan die van de originele PGA-notatie voor element x tijdens taak y.
- $H_{\text{efficiëntie}}$: de interpretatieve efficiëntie van de nieuwe PGA-notatie voor element x tijdens taak Y is hoger dan die van de originele PGA-notatie voor element x tijdens taak y.
- $H_{\text{representatievoorkeur}}$: de representatievoorkeur van de nieuwe PGA-notatie voor element x is hoger dan die van de originele PGA-notatie voor element x tijdens taak y.
- $H_{\text{geaggregeerd}}$: de geaggregeerde score van de nieuwe PGA-notatie voor element x is hoger dan die van de originele PGA-notatie voor element x.

Aan het experiment hebben 14 participanten en 44 participanten meegedaan. Er is gewerkt met een dubbele analyse om de resultaten met elkaar te kunnen vergelijken. In relatie tot de overkoepelende onderzoeksvraag is er gekozen voor het uitvoeren van een experiment binnen een praktische bedrijfscontext door op basis van de literatuur een experimentele methode op te stellen om de intuïtiviteit tussen twee verschillende versies PGA-modelleertaal te evalueren. Daaruit blijkt dat de oorspronkelijke methodes niet voorzien in het meten van de intuïtiviteit tussen twee verschillende versies van conceptuele modelleertalen en deze te evalueren. Het experiment bestaat uit het beantwoorden van contextvragen, notatie-associatievragen, begripsvragen, herinneringsvragen en representatievoorkeursvragen om de hypothesen te toetsen. De PGA-elementen die in het onderzoek zijn onderzocht zijn: *customer goal, value proposition, competence, internal goal* en *value stream*.

De uitkomsten van de metingen van interpretatieve effectiviteit en representatievoorkeur zijn geaggregeerd per PGA-element. Door deze resultaten samen te voegen voor de taken notatie-associatievragen, begripsvragen, herinneringsvragen en representatievoorkeursvragen ontstaat er een intuïtieve voorkeur per PGA-element. Op basis van de resultaten van dit onderzoek is gebleken dat het niet is aangetoond dat de nieuwe versie van de PGA-notatie intuïtiever is dan de oorspronkelijke PGA-notatie. Ook is er onderzoek gedaan naar de efficiëntie van het beantwoorden van vragen. De resultaten over de taken heen hebben niet aangetoond dat het beantwoorden van

vragen efficiënter is voor de nieuwe versie van de PGA-notatie dan voor de oorspronkelijke PGA-notatie.

De resultaten van de representatievoorkeursvragen hebben aangetoond dat voor de PGA-elementen *customer goal* en *value proposition* de voorkeur is voor de nieuwe PGA-notatie, maar in combinatie met de notatie-associatievragen, begripsvragen en herinneringsvragen is dit niet significant gebleken. De resultaten en de bijbehorende toelichting bij representatievoorkeursvragen geven aanleiding om de PGA-elementen te herontwerpen door een expert op het gebied van grafische vormgeving. Deze nieuwe PGA-elementen kunnen vervolgens weer geëvalueerd worden op basis van het onderzoek van Roelens en Bork (2020). Vervolgens kan op basis van de onderzoeksmethode, beschreven in dit onderzoek de nieuwe PGA-elementen worden geëvalueerd met de oorspronkelijke versie.

Ook is op te merken dat door het uitvoeren van eerst notatie-associatievragen, begripsvragen, herinneringsvragen en vervolgens de representatievoorkeursvragen er een zekere mate van voorkennis ontstaan is, wanneer naar de volgende taak wordt gegaan. Dit heeft mogelijk invloed op de intuïtieve voorkeur. Een aanbeveling voor verder onderzoek zou kunnen zijn om te onderzoeken wat het effect is op de intuïtieve begrijpelijkheid om representatievoorkeursvragen te combineren met notatie-associatievragen, begripsvragen en herinneringsvragen.

Summary

This research aims to evaluate the intuitiveness of two different versions of Process Goal Alignment (PGA) notation within a practical business context. Bork et al. (2019) describe an evaluation technique to assess the intuitiveness of a modeling language. This evaluation technique has been successfully applied by Roelens and Bork (2020), resulting in a number of proposed improvements to the current PGA notation. This research is about determining whether these improvements to the intuitiveness of PGA notation are also significant for potential end users.

The problem definition is: "The intuitiveness of the new version of the PGA notation has not been evaluated by conducting the experiment within a practical business context".

The research question is: "How can the intuitiveness of two different versions of the PGA notation be evaluated by conducting an experiment within a practical business context?"

This experiment examined whether the independent variables (i.e., current PGA notation, new PGA notation) can influence the dependent (i.e., interpretative effectiveness, interpretative efficiency and intuitive comprehensibility). In order to measure the dependent variables, the following hypotheses have been developed:

- H_effectiveness: the interpretative effectiveness of the new PGA notation is higher than that of the original PGA notation.
- H_efficiency: the interpretative efficiency of the new PGA notation is higher than that of the original PGA notation.
- H_representation preference: the representation preference of the new PGA notation for element x is higher than that of the original PGA notation for element x during task y.
- H_aggregate: the aggregate score of the new PGA notation for element x is higher than that of the original PGA notation for element x.

14 participants and 44 participants participated in the experiment. A double analysis was used to compare the results. In relation to the overall research question, the choice was made to conduct an experiment within a practical business context by establishing an experimental method based on the literature to evaluate intuitiveness between two different versions of PGA modeling language. This shows that the original methods do not provide for measuring and evaluating intuitiveness between two different versions of conceptual modeling languages. The experiment consists of answering context questions, notation association questions, concept questions, reminder questions and representation preferred questions to test the hypotheses. The PGA elements investigated in this research are customer goal, value proposition, competence, internal goal and value stream.

The results of the measures of effectiveness and representation preference were aggregated by PGA element. Combining these results for the tasks notation association questions, comprehension questions, reminder questions, and representation preference questions creates intuitive preference per PGA element. Based on the results of this study, it was found that it has not been shown that the new version of the PGA notation is more intuitive than the original PGA notation. The efficiency of answering questions was also investigated. The results across tasks have not shown that answering questions is more efficient for the new version of the PGA notation than for the original PGA notation.

The results of the representation preference questions showed that for the PGA elements customer goal and value proposition there is a preference for the new PGA notation, but in combination with the notation association questions, comprehension questions and reminder questions this was not

found to be significant. The results and associated explanations for representation preference questions prompted a redesign of the PGA elements by an expert in graphic design. These revised PGA elements can then be evaluated again based on the research of Roelens and Bork (2020). Then, based on the research method described in this study, the revised PGA elements can be evaluated with the original version. It can also be noticed that by performing first the notation association questions, comprehension questions, recall questions and then the representation preference questions there is a certain amount of prior knowledge created, when moving to the next task. This may have an impact on intuitive comprehension. A recommendation for further research might be to investigate the effect on intuitive comprehensibility of combining representation preference questions with notation association questions, comprehension questions, and recall questions.

Inhoudsopgave

Abstract	ii
Sleutelbegrippen	ii
Samenvatting	iii
Summary	v
Inhoudsopgave	vii
1. Introductie	1
1.1. <i>Achtergrond</i>	1
1.2. <i>Gebiedsverkenning</i>	1
1.3. <i>Probleemstelling</i>	2
1.4. <i>Opdrachtformulering</i>	2
1.5. <i>Motivatie/relevantie</i>	3
1.6. <i>Aanpak in hoofdlijnen</i>	4
2. Theoretisch kader	5
2.1. <i>Onderzoeks aanpak</i>	5
2.1.1. <i>Zoekstrategie</i>	5
2.1.2. <i>Zoekcriteria voor artikelen</i>	5
2.1.3. <i>Selectiecriteria</i>	5
2.1.4. <i>Stopcriteria</i>	6
2.2. <i>Uitvoering</i>	6
2.2.1. <i>Building blocks methode</i>	6
2.2.2. <i>Forward snowballing methode</i>	7
2.3. <i>Resultaten en conclusies</i>	7
2.3.1. <i>Resultaten</i>	7
2.3.2. <i>Conclusies</i>	10
2.4. <i>Doel van het vervolgonderzoek</i>	11
3. Methodologie	12
3.1. <i>Conceptueel ontwerp: keuze van onderzoeksmethode(n)</i>	12
3.2. <i>Technisch ontwerp: uitwerking van de methode</i>	13
3.3. <i>Gegevensanalyse</i>	17
3.4. <i>Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten</i>	18
4. Resultaten	20
4.1. <i>Uitvoering van het onderzoek</i>	20
4.2. <i>Uitkomsten: effectiviteit</i>	20
4.3. <i>Uitkomsten: efficiëntie</i>	22
4.4. <i>Uitkomsten: representatievoorkeur</i>	25

4.5.	<i>Uitkomsten: geaggregeerd per PGA-element</i>	27
4.6.	<i>Conclusie</i>	28
5.	Discussie, conclusies en aanbevelingen	30
5.1.	<i>Discussie – reflectie</i>	30
5.2.	<i>Conclusies</i>	31
5.3.	<i>Aanbevelingen voor de praktijk</i>	32
5.4.	<i>Aanbevelingen voor verder onderzoek</i>	32
	Referenties	33
	Bijlage A: Resultaten building blocks methode	35
	Bijlage B: Resultaten forward snowballing methode	36
	Bijlage C: Samenvatting literatuuronderzoek	37
	<i>H1 Artikel: Burton-Jones et al. (2009)</i>	37
	<i>H2 Artikel: Cruzes et al. (2013)</i>	38
	<i>H3 Artikel: Gemino en Wand (2004)</i>	39
	<i>H4 Artikel: Mendling et al. (2018)</i>	40
	<i>H5 Artikel: Recker et al. (2011)</i>	41
	<i>H6 Artikel: Shanks et al. (2008)</i>	42
	Bijlage D: Survey vragen	44
	<i>H1 Datasets en export van de vragen</i>	44
	<i>H2 Codelijst van de vragen</i>	44
	Bijlage E: Persoonlijke resultaten van de notatie-associatievragen	46
	<i>H1 Normale verdeling</i>	46
	<i>H2 Mann-Whitney U test</i>	47
	<i>H3 Independent Samples Test</i>	48
	Bijlage F: Groepsresultaten van de notatie-associatievragen	51
	<i>H1 Normale verdeling</i>	51
	<i>H2 Mann-Whitney U test</i>	52
	Bijlage G: Persoonlijke resultaten van de begripsvragen	54
	<i>H1 Normaalverdeling</i>	54
	<i>H2 Mann-Whitney U test</i>	54
	<i>H3 Independent Samples Test</i>	56
	Bijlage H: Groepsresultaten van de begripsvragen	58
	<i>H1 Normaleverdeling</i>	58
	<i>H2 Mann-Whitney U test</i>	58
	Bijlage I: Persoonlijke resultaten van herinneringsvragen	61

<i>H1 Normale verdeling</i>	61
<i>H2 Mann-Whitney U test</i>	61
<i>H3 Independent Samples Test</i>	63
Bijlage J: Groepsresultaten van herinneringsvragen	65
<i>H1 Normale verdeling</i>	65
<i>H2 Mann-Whitney U test</i>	65
Bijlage K: Persoonlijke resultaten van de Representatievoorkeursvragen	68
<i>H1 Normale verdeling</i>	68
<i>H2 One-Sample Test</i>	68
Bijlage L: Groepsresultaten van de representatievoorkeursvragen	72
<i>H1 Normale verdeling</i>	72
<i>H2 One-Sample Test</i>	72
<i>H4 resultanten toelichting op de representatievoorkeursvragen</i>	76
Bijlage M: Persoonlijke resultaten van de elementen geaggregeerd	81
<i>H1 Normale verdeling</i>	81
<i>H2 Mann-Whitney U test</i>	81
<i>H3 Independent Samples Test</i>	82
Bijlage N: Groepsresultaten van de elementen geaggregeerd	84
<i>H1 Normale verdeling</i>	84
<i>H2 Mann-Whitney U test</i>	84
<i>H3 Independent Samples Test</i>	85
Bijlage O: Analyseprocedure	87

1. Introductie

Dit hoofdstuk beschrijft waarom het nodig is om de *Process Goal Alignment* (PGA) notatie (Roelens et al., 2019) te evalueren en hoe de uitkomsten van het experiment kunnen bijdragen aan het verder ontwerpen van de experimentele evaluatietechniek (Bork et al., 2019).

1.1. Achtergrond

Enterprise architectuur (EA) *frameworks* zijn “een samenhangend geheel van principes, methodes en modellen die gebruikt worden bij het ontwerp en de realisatie van de organisatiestructuur, bedrijfsprocessen, informatiesystemen en technische infrastructuren van een onderneming” (Lankhorst, 2017, p. 3). Lankhorst (2017) beschrijft dat architectuurmodellen niet alleen nuttig zijn om inzicht te geven in de oorspronkelijke of toekomstige situatie, maar ook gebruikt worden bij analyses om de oorspronkelijke en toekomstige situatie te evalueren. Twee bekende EA *frameworks* zijn Zachman (Zachman, 1982) en TOGAF (i.e. The Open Group Architecture Framework) (OpenGroup, 2018). EA producten zijn “een samenhangend geheel van diensten en/of passieve structuurelementen, voorzien van een contract/set van afspraken, dat in zijn geheel wordt geleverd aan interne of externe klanten” (Lankhorst, 2017, p. 95). EA producten worden in een modelleertaal vastgelegd. Een modelleertaal omvat de taalsyntax (i.e. de grammatica van de taal), de taalsemantiek (i.e. de betekenis van de taalconcepten) en de taalnotatie (i.e. de visuele representatie van de taal) (Bork et al., 2019).

Modelleertalen kunnen globaal onderverdeeld worden als domeinspecifieke modelleertaal of als generieke modelleertaal. Een domeinspecifieke modelleertaal is “een op taal gebaseerde aanpak die het abstractieniveau verhoogt door te modelleren met behulp van domeinconcepten” (Luoma et al., 2004, p. 4). Een domeinspecifieke modelleertaal wordt gebruikt om concepten in kaart te brengen binnen de context van een specifiek domein (Roelens & Bork, 2020). Een generieke modelleertaal bestaat uit domeinonafhankelijke concepten, bijvoorbeeld Archimate (OpenGroup, 2019) en BPMN (i.e. *Business Process Model and Notation*) (OMG, 2014). Zo is Archimate een generieke modelleertaal voor het beschrijven van de bedrijfsarchitectuur, applicatie-architectuur en technologie-architectuur (OpenGroup, 2019) en BPMN een modelleertaal om processen vast te leggen.

Bij modelleertalen speelt de visuele weergave of notatie van elementen een belangrijke rol. Een intuïtieve notatie is belangrijk voor het correct gebruik en begrip van het model. Jošt et al. (2016, p. 91) beschrijven intuïtieve begrijpelijkheid als “het gemak waarmee het diagram zonder enige voorkennis of opleiding onmiddellijk door de gebruikers kan worden begrepen”.

Een intuïtieve notatie betekent dat eindgebruikers de betekenis kunnen afleiden vanuit de visuele weergave, wat ervoor zorgt dat de cognitieve belasting om de modelleertaal te begrijpen en te gebruiken aanzienlijk verminderd wordt (Bork et al., 2019).

1.2. Gebiedsverkenning

De PGA-modelleertaal is een domeinspecifieke modelleertaal binnen EA. Roelens et al. (2019) ontwierp de PGA techniek als een begrijpelijke modelleertaal die verschillende mechanismen combineert tot een totaalaanpak om strategische afstemming te realiseren binnen de bedrijfsarchitectuur van een onderneming. Strategische afstemming houdt in dat de strategische positionering van de onderneming wordt afgestemd op het ontwerp van de activiteiten die deze strategie ondersteunen. PGA bestaat uit een modelleertaal met concepten uit strategische management raamwerken, een systeem voor het instellen en meten van prestaties en een *heat mapping* weergave op basis van het prestatie-meetsysteem aangevuld met een prioriteitsmechanisme (Roelens et al., 2019).

Bork et al. (2019) beschrijven een evaluatietechniek om de intuïtiviteit van een modelleertaal te beoordelen. Deze evaluatietechniek bestaat uit drie fases, die elk een specifiek experimentele taak met de participanten inhouden. Voorafgaand aan het experiment vindt er een initiatiefase plaats en na het uitvoeren van het experiment wordt in de afsluitende fase kwantitatieve en kwalitatieve feedback gevraagd over het experiment.

In de *begrip associatie* fase krijgen participanten een lijst van begrippen aangeleverd die verwijzen naar de naam van een modelleertaalconcept. De participanten maken individueel een grafische weergave van het begrip. De resultaten worden geanalyseerd op vorm, kleur en andere visuele variabelen, zoals iconen, tekst en grafische plaatsing (Moody, 2010). De vergelijking van de grafische weergaven met de bestaande notatie van het begrip zou een tekortkoming kunnen identificeren en een verbetervoorstel kunnen opleveren.

In de *notatie associatie* fase krijgen participanten de bestaande visuele notaties te zien van de modelleertaal. Vervolgens wordt aan de participanten gevraagd om maximaal drie associaties te noteren bij het bekijken van de visuele notaties. Als een van de notaties overeenkomt met de werkelijke naam of betekenis dan is de notatie volledig geïdentificeerd. Wanneer de genoemde associatie bijna overeenkomt met de notatie dan wordt deze geclassificeerd als gedeeltelijk geïdentificeerd. Als er geen overeenkomst is met de werkelijke naam of betekenis dan wordt de notatie geclassificeerd als niet geïdentificeerd.

De *casestudie fase* is gericht op het testen of participanten in staat zijn om modelleertaalconcepten intuïtief te combineren. Om dit te testen zijn er twee mogelijkheden: door participanten zelf een bepaalde casestudie te laten modelleren of door het beantwoorden van begripsvragen. Dit laatste wordt gezien als een alternatief als de tijd beperkt is of als het aantal participanten hoog is. Roelens en Bork (2020) gebruikten de laatste variant bij de evaluatie van de PGA-notatie.

1.3. Probleemstelling

Om de intuïtiviteit van de PGA-notatie te beoordelen is de experimentele evaluatietechniek van Bork et al. (2019) toegepast. Deze evaluatietechniek is succesvol toegepast op de PGA-modelleertaal (Roelens & Bork, 2020), wat geleid heeft tot een aantal voorgestelde verbeteringen aan de oorspronkelijke PGA-notatie. Het is echter niet vast te stellen of deze verbeteringen aan de intuïtiviteit van de notatie ook significant zijn voor potentiële eindgebruikers. Er heeft immers nog geen evaluatie plaatsgevonden om de intuïtiviteit van de nieuwe PGA-notatie ook effectief aan te tonen. Er bestaan dus twee verschillende versies van de PGA-notatie die met elkaar vergeleken moeten worden. Hiervoor is verder experimenteel onderzoek noodzakelijk.

De probleemstelling is als volgt geformuleerd:

“De intuïtiviteit van de nieuwe versie van de PGA-notatie is niet geëvalueerd door het uitvoeren van het experiment binnen een praktische bedrijfscontext”.

1.4. Opdrachtformulering

Het doel van het onderzoek is om een experimenteel onderzoek op te zetten en uit te werken om de intuïtiviteit van twee verschillende versies van de PGA-notatie te evalueren binnen een praktische bedrijfscontext.

De overkoepelende onderzoeksvraag is als volgt geformuleerd:

Hoe kan de intuïtiviteit van twee verschillende versies van de PGA-notatie geëvalueerd worden door het uitvoeren van een experiment binnen een praktische bedrijfscontext?

Om de onderzoeksvraag te kunnen beantwoorden is een gestructureerde aanpak nodig. Stap 1 is het uitvoeren van een literatuuronderzoek om methodes te identificeren die relevant zijn om de conceptuele modelleertalen empirisch te evalueren. In stap 2 wordt op basis van de uitkomsten uit het literatuuronderzoek in stap 1 een experimentele methode uitgewerkt. Deze experimentele methode beschrijft waarom het experiment kan bijdragen aan het evalueren van de intuïtiviteit van de PGA-modelleertaal en hoe het onderzoek op een verantwoorde manier wordt opgezet. In stap 3 wordt vervolgens de experimentele methode toegepast door het uitvoeren van een experiment binnen een praktische bedrijfscontext. De resultaten worden vastgelegd en geëvalueerd met als doel om aan te tonen hoe intuïtief de nieuwe PGA-notatie is in vergelijking met de vorige versie. Voor elke stap is een deelvraag geformuleerd. De antwoorden op deze deelvragen dragen dus bij aan het beantwoorden van de overkoepelende onderzoeksvraag.

De deelvragen zijn als volgt geformuleerd:

1. Welke methodes zijn relevant om conceptuele modelleertalen empirisch te evalueren?
2. Welke methodologische elementen kunnen bijdragen aan het evalueren van de intuïtiviteit van de PGA-modelleertaal?
3. Hoe intuïtief is de nieuwe PGA-notatie in vergelijking met de vorige versie?

1.5. Motivatie/relevantie

Het voordeel van de PGA-modelleertaal als domeinspecifieke modelleertaal is dat deze gericht is op het afstemmen van de strategie en de processen door het beoordelen en intuïtief visualiseren van de onderliggende samenhang. In vergelijking met een generieke modelleertaal maakt een domeinspecifieke modelleertaal het mogelijk om de complexiteit te verminderen door gebruik te maken van concepten die bekend zijn bij de beoogde eindgebruikers (Roelens & Bork, 2020). Strategische afstemming binnen de bedrijfsarchitectuur is een belangrijk aandachtspunt binnen de onderneming, omdat de bedrijfsstrategie wordt afgestemd op de interne infrastructuur en processen (Roelens & Bork, 2020). De PGA-modelleertaal is gericht op de bedrijfsarchitectuur van een onderneming (Roelens et al., 2019). “ Bedrijfsarchitectuur is het vakgebied dat – gegeven de strategische keuzes van het topmanagement – zorgt voor een integrale en samenhangende bedrijfsinrichting” (Schraesande, 2010). De essentie van bedrijfsarchitectuur wordt weergegeven in het artikel van Oberconsulting (2020) en geeft de volgende beschrijving “Met een bedrijfsarchitectuur ondersteunen we de realisatie van strategische doelen door een hierop aansluitende inrichting en werking in beeld te brengen”. Dit zorgt voor standaardisatie van de bedrijfsprocessen die voldoen aan de strategische operationele modellen van de organisatie en vormen zo een link tussen de bedrijfsstrategie en de implementatie of uitvoering ervan (Realdolmen, 2016).

Het doel is om een experimenteel onderzoek op te zetten en uit te voeren om de intuïtiviteit van de PGA-modelleertaal te evalueren. Dit is nodig omdat in de oorspronkelijke wetenschappelijke literatuur een duidelijk gebrek bestaat aan dergelijke experimentele evaluatiemethodes in relatie tot de intuïtiviteit van conceptuele modelleringstechnieken. Het onderzoek draagt bij aan het evalueren van de nieuwe versie van de PGA-notatie door het uitvoeren van een experiment binnen een praktische bedrijfscontext.

1.6. Aanpak in hoofdlijnen

Dit afstudeerrapport is als volgt opgebouwd:

- Hoofdstuk 1 beschrijft waarom het nodig is om de intuïtiviteit van de PGA-notatie vergelijkend te evalueren en hoe de uitkomsten van het experiment kunnen bijdragen aan het ontwerp van een experimentele evaluatie techniek.
- Hoofdstuk 2 beschrijft het theoretisch kader, de onderzoeks aanpak en de resultaten van de literatuurstudie. Vervolgens worden methodes geïdentificeerd die relevant zijn om de conceptuele modelleertalen empirisch te kunnen evalueren (i.e. deelonderzoeksvraag 1).
- Hoofdstuk 3 beschrijft de experimentele methode en waarom het bijdraagt aan het evalueren van de intuïtiviteit van de PGA-modelleertaal en hoe het onderzoek op een verantwoorde manier is opgezet (i.e. deelonderzoeksvraag 2).
- Hoofdstuk 4 beschrijft de uitvoering en de resultaten van het onderzoek gebaseerd op de experimentele methode die ontwikkeld is op basis van de geselecteerde literatuur. (i.e. deelonderzoeksvraag 3).
- Hoofdstuk 5 geeft antwoord op de overkoepelende onderzoeksvraag, de verklaring van het onderzoek en geeft aanbevelingen voor verder onderzoek.

2. Theoretisch kader

Dit hoofdstuk beschrijft het theoretisch kader, de onderzoeksaanpak, uitvoering en resultaten van de literatuurstudie.

2.1. Onderzoeksaanpak

Het doel van het literatuuronderzoek is om te achterhalen welke methodes relevant zijn om conceptuele modelleertalen empirisch te evalueren (i.e. deelonderzoeksvraag 1). De geselecteerde artikelen dienen een bijdrage te leveren om kennis op te doen over het empirisch evalueren van conceptuele modellen.

2.1.1. Zoekstrategie

Als zoektermen zijn “*conceptual modeling*” en “*empirical evaluation*” gebruikt in de *building blocks* methode, waarbij een ‘AND’ werd gebruikt om de termen te combineren in het zoekscherm van de OU bibliotheek. De zoektermen moeten dus allebei terugkomen in het artikel. Deze zoektermen zijn afgeleid van deelonderzoeksvraag 1. Het resultaat is dat de gevonden artikelen gaan over het empirisch evalueren van conceptuele modellen.

Daarnaast is de zoekstrategie *forward snowballing* toegepast. Het artikel van Gemino en Wand (2004) is als uitgangspunt genomen om te zoeken naar vervolgp publicaties die voortbouwen op deze publicatie door op *cited by* te selecteren bij het artikel van Gemino en Wand (2004) in de Open Universiteit bibliotheek.

2.1.2. Zoekcriteria voor artikelen

De zoekcriteria voor het vinden van artikelen bestaan uit de volgende vijf criteria:

- Database: Open Universiteit bibliotheek is gebruikt als database. Deze bron biedt een set aan artikelen die volledig toegankelijk zijn zonder extra kosten.
- Publicatiedatum: publicatiedatum is na 2004. Het werk van Gemino en Wand (2004) is immers als basisartikel gekozen voor de literatuurstudie.
- Taalkeuze: Engels. Dit is de standaardtaal in het wetenschappelijk domein.
- Krantenartikelen en *book reviews* zijn uitgesloten in de zoekresultaten. Hiermee beperkt het zoekresultaat zicht tot conferentiepublicaties, artikelen in wetenschappelijke tijdschriften en boeken.
- Volledig tekst online. De artikelen zijn volledig beschikbaar en kunnen geraadpleegd worden voor de literatuurstudie.

De bovenstaande zoekcriteria worden automatisch meegegeven in het zoekalgoritme, zodat dit niet handmatig gecontroleerd hoeft te worden.

2.1.3. Selectiecriteria

Om te beslissen of een artikel relevant is, wordt een *quick scan* gedaan per artikel. De titel, *abstract*, inleiding en conclusie worden volledig doorgelezen en getoetst aan de inhoudelijke aspecten: context, respondenten en type vraag die in tabel 1 zijn weergegeven. Daarnaast wordt er d.m.v. zoeken op steekwoorden gekeken in het artikel om de elementen ‘respondenten’ en ‘typevraag’ inhoudelijk te kunnen beoordelen. Voor respondenten is gezocht op *business*, *student* en *respondent* en voor typevraag op *question*, *comprehension*, *problem*, *close* en *knowledge*.

Voor de *building blocks methode* geldt dat geschikte literatuur inhoudelijk moet gaan over empirisch evalueren van conceptuele modellen en is uitgevoerd door participanten die werken bij een organisatie.

Voor de forward *snowballing* methode geldt dat geschikte literatuur een duidelijke relatie moet hebben met de uitgangspunten van het empirisch evalueren van conceptuele modellen en dat het experiment is uitgevoerd door participanten die werken bij een organisatie. Het zoeken door middel van *forward snowballing* levert ook artikelen op met alleen een referentie naar het artikel van Gemino en Wand (2004). Deze artikelen bouwen inhoudelijk niet voort op de ideeën uit het artikel en zijn daarmee niet geschikt als bron voor de literatuurstudie.

Tabel 1: Inhoudelijke aspecten op basis van meta-data, titel & abstract, inleiding en conclusie

Element	Omschrijving	Selectiecriteria
Context	Zoektermen “ <i>conceptual modeling</i> ” en “ <i>empirical evaluation</i> ” komen voor in het artikel.	Het artikel moet inhoudelijk gaan over empirisch evalueren van conceptuele modellen.
Respondenten	Het experiment kan uitgevoerd zijn door studenten en/of business expert.	Het experiment is door business experts uitgevoerd.
Type vraag	Gemino en Wand (2004) beschrijven de onderstaande vijf type vragen. <i>Recal:</i> Herinneringsvragen. Comprehension: Begripsvragen gaan over elementen en informatie die in het model terug te vinden zijn. Problem-solving: Probleemoplossende taken waarbij de participanten moeten redeneren over het domein dat wordt gemodelleerd. Cloze: Het houden van een invulproef. De participanten worden gevraagd om de blanco 's in een paragraaf in te vullen nadat het model is verwijderd. Knowledge: Kennisvragen over procesmodelleringsconcepten om het kennisniveau van de participanten in kaart te brengen in de vorm van een kennistoets, zodat hier op geanticipeerd kan worden bij het stellen van begripsvragen.	Het artikel beschrijft één of meerdere type vragen.

2.1.4. Stopcriteria

Het stopcriterium treedt binnen een zoekopdracht in werking als er drie artikelen niet voldoen aan het ‘context’ selectiecriteria (zie §2.1.3).

2.2. Uitvoering

2.2.1. Building blocks methode

Als zoektermen zijn “*conceptual modeling*” en “*empirical evaluation*” als zoekquery ingevuld en dit resulteert in 180 resultaten gesorteerd op relevantie. Na inhoudelijke analyse van de 12 artikelen (zie bijlage A) is het stopcriterium (zie §2.1.4) in werking getreden. De artikelen die voldoen aan het ‘context’ selectiecriteria (zie §2.1.3) zijn opgenomen in tabel 2.

Tabel 2: Geselecteerde artikelen building blocks methode

Referentie	Titel	Jaar	Context	Respondent	Type vraag
Burton-Jones et al. (2009)	<i>Guidelines for Empirical Evaluations of Conceptual Modeling Grammars</i>	2009	Ja	Business expert, studenten	Comprehension, Problem-solving
Cruzes et al. (2013)	<i>Empirical Evaluation of the Quality of Conceptual Models Based on User Perceptions: A Case Study in the Transport Domain</i>	2013	Ja	Business expert	Comprehension
Gemino en Wand (2004)	<i>A framework for empirical evaluation of conceptual modeling techniques</i>	2004	Ja	Business expert	Recal, Comprehension, Problem-solving, Cloze
Recker et al. (2011)	<i>Do Ontological Deficiencies in Modeling Grammars Matter?</i>	2011	Ja	Business expert	Comprehension
Shanks et al. (2008)	<i>Representing Part-Whole Relations in Conceptual Modeling: An Empirical Evaluation</i>	2008	Ja	Business expert	Problem-solving

2.2.2. Forward snowballing methode

Voor de *forward snowballing* methode is het artikel van Gemino en Wand (2004) gekozen en daarbij zijn 33 resultaten weergegeven die een referentie naar Gemino en Wand (2004) hebben opgenomen. Deze lijst is opgenomen in bijlage B. In totaal zijn er negen artikelen inhoudelijk bekeken. Van drie artikelen kon de volledige tekst niet *gedownload* worden en zijn daarom uitgesloten. Na negen artikelen is het stopcriterium in werking getreden. Het artikel dat voldoet aan de selectiecriteria (zie §2.1.3) is opgenomen in tabel 3.

Tabel 3: Geselecteerde artikelen forward snowballing methode

Referentie	Titel	Jaar	Context	Respondent	Type vraag
Mendling et al. (2018)	<i>An Empirical Review of the Connection Between Model Viewer Characteristics and the Comprehension of Conceptual Process Models</i>	2018	Ja	Business expert	Comprehension, Knowledge

2.3. Resultaten en conclusies

2.3.1. Resultaten

Deze paragraaf beschrijft de resultaten van het literatuuronderzoek om te achterhalen welke methodes zijn toegepast om conceptuele modelleertalen te evalueren (i.e. deelonderzoeksvraag 1). Per onderzochte referentie (zie tabel 2 en tabel 3) zijn de volgende vijf concepten onderzocht tijdens de literatuurstudie:

- Onderzoeksopzet: de onderzoeksopzet dat toegepast is.
- Aantal participanten: het aantal participanten dat heeft meegedaan aan het onderzoek.
- De onafhankelijke variabelen: “variabelen die worden gemanipuleerd of gewijzigd om de impact op een afhankelijke variabele te meten” Saunders et al. (2019, p. 805).
- Afhankelijke variabelen: “variabelen die kunnen veranderen als gevolg van veranderingen in andere variabelen” Saunders et al. (2019, p. 801).
- Metingen: metingen om de onafhankelijke, afhankelijke variabelen te toetsen.

De resultaten staan hieronder kort beschreven. De samenvattingen van de referenties zijn opgenomen in bijlage C.

Onderzoeksopzet

Gemino en Wand (2004) presenteren een *framework* voor het empirisch evalueren van conceptuele modelleringstechnieken en Burton-Jones et al. (2009) beschrijven richtlijnen om conceptuele modellen empirisch te evalueren. Cruzes et al. (2013) hebben gekozen voor een semigestructureerd interview om de kwaliteit van een conceptueel model te evalueren in de transportlogistiek, waarbij gebruik gemaakt wordt van bestaande evaluatiemethodes. Het artikel bouwt met name voort op de vragen en metingen van Maes en Poels (2007).

Shanks et al. (2008), Recker et al. (2011) en Mendling et al. (2018) hebben gekozen voor een experiment/survey. Shanks et al. (2008) hebben onderzoek gedaan naar theorieën over alternatieve notaties van conceptuele modellering. Mendling et al. (2018) hebben onderzocht welke factoren van invloed zijn op de manier waarop persoonlijke kenmerken het begrip van conceptuele modellen beïnvloeden, wat belangrijk is voor het verbeteren van het ontwerp van informatiesystemen. Recker et al. (2011) beschrijven in het artikel hoe de eigenschappen van een conceptuele modelleringsgrammatica aansluiten bij het waargenomen nut en het waargenomen gebruikersgemak.

Aantal participanten

Aan het onderzoek van Gemino en Wand (2004) en Burton-Jones et al. (2009) hebben geen participanten meegedaan, omdat het een onderzoek betrof rondom het opstellen van een *framework* of richtlijnen. Aan het onderzoek van Cruzes et al. (2013) hebben zeven participanten meegedaan en aan het onderzoek van Shanks et al. (2008) 30 participanten. Tenslotte hebben 590 participanten meegedaan aan het onderzoek van Recker et al. (2011).

Onafhankelijke variabelen

In tabel 4 is een overzicht weergegeven van de onafhankelijke variabelen die geselecteerd zijn uit de referentie artikelen. De onafhankelijke variabelen gaan over de syntax/betekenis van de grammatica (Burton-Jones et al., 2009), het begrijpen van het conceptuele model (Cruzes et al., 2013; Gemino & Wand, 2004; Shanks et al., 2008), waargenomen tekort (Recker et al., 2011) en de demografische factoren (Mendling et al., 2018).

Tabel 4: Overzicht van de onafhankelijke variabelen

Referentie	Onafhankelijke variabele	Omschrijving
Burton-Jones et al. (2009)	1. <i>Syntax</i> 2. <i>Semantics</i> 3. <i>Pragmatics</i>	1. De concepten in de grammatica en de regels voor de indeling. 2. De betekenis van de concepten in de grammatica. 3. De context waarin de grammatica wordt gebruikt.
Cruzes et al. (2013)	1. Begrijpelijkheid	1. Overtuiging en het gemak van het succes van het model.
Gemino en Wand (2004)	1. Begrijpelijkheid	1. Interpreteren van het conceptuele model.
Mendling et al. (2018)	1. Positie 2. Theorie 3. Modelleerjaren ervaring: 4. Modellen gelezen 5. Modellen gecreëerd 6. Training 7. Zelfstudie 8. Niveau 9. Maanden BPNM	1. Het vastleggen van de positie van de participant in relatie tot het onderzoek: student, practitioner of anders. 2. Niveau van theoretische kennis van modellering. 3. Aantal jaren ervaring in modellering. 4. Aantal modellen gelezen/geanalyseerd in de afgelopen 12 maanden. 5. Aantal modellen gecreëerd of aangepast in de afgelopen 12 maanden. 6. Aantal dagen modelleer trainingen gevolgd in de afgelopen 12 maanden. 7. Aantal dagen zelfstudie gevolgd in de afgelopen 12 maanden. 8. Niveau van bekendheid met een specifieke modelleringsgrammatica. 9. Aantal maanden werkervaring met een specifieke modelleringsgrammatica
Recker et al. (2011)	1. <i>Perceived construct deficit</i>	1. Het waargenomen tekort in het actualiseren van logs en statusveranderingen.

Referentie	Onafhankelijke variabele	Omschrijving
	<ol style="list-style-type: none"> 2. <i>Perceived construct redundancy</i> 3. <i>Perceived construct overload</i> 4. <i>Perceived construct excess</i> 	<ol style="list-style-type: none"> 2. De waargenomen redundantie van BPMN grammatica in de verwoording van echte objecten. 3. Waargenomen overbelasting. 4. Waarnemingen van constructiefouten.
Shanks et al. (2008)	<ol style="list-style-type: none"> 1. Begrijpelijkheid 	<ol style="list-style-type: none"> 1. Probleemoplossend vermogen.

Afhankelijke variabelen

In tabel 5 is een overzicht weergegeven van de afhankelijke variabelen die geselecteerd zijn uit de referentieartikelen die gaan over het waargenomen gemak van het gebruik van het model. In de referentieartikelen gaan de afhankelijke variabelen over betrouwbaarheid (Burton-Jones et al., 2009), begrip van het model, (Burton-Jones et al., 2009; Cruzes et al., 2013; Gemino & Wand, 2004; Mendling et al., 2018; Recker et al., 2011), ingezette middelen (Burton-Jones et al., 2009), kwaliteit (Cruzes et al., 2013), doeltreffendheid (Gemino & Wand, 2004), efficiëntie (Gemino & Wand, 2004; Mendling et al., 2018; Shanks et al., 2008) en voltooijsheid (Mendling et al., 2018).

Tabel 5: Overzicht van de afhankelijke variabelen

Referentie	Afhankelijke variabele	Omschrijving
Burton-Jones et al. (2009)	<ol style="list-style-type: none"> 1. <i>Representational fidelity</i> 2. <i>Representational efficiency</i> 3. <i>Interpretational fidelity</i> 4. <i>Interpretational efficiency</i> 	<ol style="list-style-type: none"> 1. Hoe vertegenwoordigt iemand de perceptie van het model of de manier waarop een groep de betekenis van het domein heeft besproken? 2. Welke middelen worden gebruikt om het model te maken? 3. Hoe betrouwbaar is de betekenis van het model? 4. Welke middelen worden gebruikt om het model te interpreteren?
Cruzes et al. (2013)	<ol style="list-style-type: none"> 1. <i>The perceived semantic quality:</i> 2. <i>The perceived ease of understanding:</i> 3. <i>The user satisfaction</i> 4. <i>The perceived usefulness</i> 	<ol style="list-style-type: none"> 1. Is een meting op hoe goed de beoogde kwaliteit wordt overgebracht en geïnterpreteerd door de ontvanger. 2. Is gebaseerd op hoe de gebruikers het model zullen beoordelen op hoe gemakkelijk het is om de informatie in het model te lezen en te kunnen begrijpen. 3. Is de inhoudelijke evaluatie van de maten waarin gebruikers tevreden zijn over het model. 4. Is een concept voor het meten van de algemene kwaliteit van een model.
Gemino en Wand (2004)	<ol style="list-style-type: none"> 1. Doeltreffendheid 2. Efficiëntie 	<ol style="list-style-type: none"> 1. Het proces waarbij individuen nadenken en communiceren over een domein met als doel om dit beter te begrijpen. 2. Welke middelen worden gebruikt om het model te interpreteren?
Mendling et al. (2018)	<ol style="list-style-type: none"> 1. <i>Performance</i> 2. <i>Voltooijsheid</i> 	<ol style="list-style-type: none"> 1. Mate van nauwkeurigheid van het begrip van het model. Dit wordt berekend door de som van de juiste antwoorden op begripsvragen die betrekken hebben op een bepaald model. 2. Deze variabele legt de tijd vast voor het voltooiën van een specifieke begrijpelijke taak.
Recker et al. (2011)	<ol style="list-style-type: none"> 1. <i>Perceived usefulness</i> 2. <i>perceived ease of use</i> 	<ol style="list-style-type: none"> 1. Waargenomen nut van de BPMN grammatica. 2. Waargenomen gebruikersgemak van de BPMN grammatica.
Shanks et al. (2008)	<ol style="list-style-type: none"> 1. <i>Performance</i> 	<ol style="list-style-type: none"> 1. Probleemoplossend vermogen, efficiëntie.

Metingen

In tabel 6 wordt het *framework* toegepast van Houy et al. (2012) voor het verder analyseren van de onderzochte referenties. In totaal zijn er zes verschillende aspecten vastgesteld: (1) herinneringsvragen, (2) begripsvragen, (3) probleemoplossing, (4) invulproef, (5) tijd nodig om het model te begrijpen en (6) waargenomen gemak van het gebruik van het model. Deze aspecten bestaan uit objectieve variabelen die gebruikt kunnen worden om het begrip van een conceptueel model te meten (1-5) en een subjectieve variabele (6).

Tabel 6: Overzicht van de gebruikte metingen

Referentie	Objectief				Subjectief	
	Doeltreffendheid				Efficiëntie	Doeltreffendheid
	1. Herinneringsvragen	2. Begripsvragen	3. Probleemoplossing	4. Invulproef (close)	5. Tijd nodig om het model te begrijpen	6. Waargenomen gemak van het gebruik van het model
Gemino en Wand (2004)	•	•	•	•	•	•
Shanks et al. (2008)			•		•	•
Recker et al. (2011)		•				•
Burton-Jones et al. (2009)		•	•			•
Cruzes et al. (2013)		•				•
Mendling et al. (2018)		•			•	•

Tabel 6 geeft aan welke type vragen gesteld kunnen worden om de doeltreffendheid in kaart te brengen en/of tijd die nodig om de efficiëntie van het model te begrijpen. De resultaten laten zien dat het waargenomen gemak van het gebruik van het model in alle referenties een belangrijk begrip is. Het aantal keren dat een type vraag is benoemd als experimentele taak is als volgt: één keer herinneringsvragen, vijf keer begripsvragen, drie keer probleemoplossende vragen en één keer een invulproef. Het meten van de tijd die nodig is om het model te begrijpen wordt drie keer toegepast.

2.3.2. Conclusies

De resultaten van het literatuuronderzoek laten zien in tabel 6 dat er verschillende methodes zijn toegepast voor het uitvoeren van het experiment. Op basis van de afhankelijke, onafhankelijke variabelen en de metingen, blijkt dat er geen eenduidige methode bestaat voor het interpreteren van het conceptuele model. Er zitten immers verschillen in de gebruikte onafhankelijke en afhankelijke variabelen en de metingen. De conclusie is dat er niet een eenduidige methode toegepast kan worden om conceptuele modelleertalen te evalueren (i.e. deelonderzoeksvraag 1). In relatie tot de overkoepelende onderzoeksvraag kan geconcludeerd worden dat de oorspronkelijke methodes nog niet voorzien in het meten van de intuïtiviteit tussen twee verschillende versies van conceptuele modelleertalen en deze te evalueren.

In dit literatuuronderzoek zijn verschillende soorten vragen geïdentificeerd om de doeltreffendheid van een conceptueel model te kunnen meten. Op basis van de resultaten van de doeltreffendheid (zie tabel 6) blijkt dat begripsvragen het meeste voorkomen en herinneringsvragen, probleemoplossende vragen, en invulvragen minder zijn toegepast.

2.4. Doel van het vervolgonderzoek

Het doel van het vervolgonderzoek is om zelf een experimenteel onderzoek op te zetten en uit te voeren om de intuïtiviteit van de PGA-modelleertaal te evalueren. Dit is nodig omdat op basis van de conclusies uit §2.3.2 blijkt dat er een gebrek bestaat aan experimentele evaluatiemethodes in relatie tot de intuïtiviteit van conceptuele modelleringstechnieken. Het doel van het vervolgexperiment is om een bijdrage te leveren aan het evalueren van de intuïtiviteit van de PGA-modelleertaal (i.e. deelonderzoekvraag 2) en om te onderzoeken hoe intuïtief de nieuwe PGA-notatie is in vergelijking met de vorige versie (i.e. deelonderzoeksvraag 3). Om deze twee deelonderzoeksvragen te kunnen beantwoorden is het noodzakelijk om de volgende activiteiten uit te voeren in H3 en H4:

- Het opzetten van een experimentele methode voor vervolgevaluatie van de intuïtiviteit van de PGA-modelleertaal.
- Het maken van keuzes voor een praktische bedrijfscontext.
- Het uitvoeren van de experimenten volgens de opgezette methode bij tiental respondenten binnen de praktische bedrijfscontext.
- Het analyseren van de intuïtiviteit van de PGA-notatie op basis van de verkregen experimentele resultaten.
- Een besluit trekken uit de evaluatie door het identificeren van eventuele benodigde aanpassingen aan de experimentele opzet.

3. Methodologie

Dit hoofdstuk beschrijft de experimentele methode en hoe dit kan bijdragen aan het evalueren van de intuïtiviteit van de PGA-modelleertaal. Tevens beschrijft dit hoofdstuk hoe het onderzoek op een verantwoorde manier is opgezet (i.e. deelonderzoeksvraag 2).

3.1. Conceptueel ontwerp: keuze van onderzoeksmethode(n)

Het doel van het onderzoek is om de intuïtiviteit van verschillende versies van de PGA-notatie te evalueren binnen een praktische bedrijfscontext. De uitdaging is om de intuïtiviteit van de nieuwe PGA-notatie aan te tonen. Om de intuïtiviteit van de PGA-modelleertaal te verbeteren is de experimentele evaluatietechniek van Bork et al. (2019) toegepast. Deze evaluatietechniek is succesvol toegepast op de PGA-modelleertaal (Roelens & Bork, 2020), wat geleid heeft tot een aantal voorgestelde verbeteringen aan de oorspronkelijke notatie. Het is echter niet vast te stellen of deze verbeteringen aan de intuïtiviteit van de notatie ook significant zijn voor potentiële eindgebruikers.

Om dit probleem op te lossen zijn informatiebronnen nodig die bestaan uit: (1) de methodes die relevant zijn om de intuïtiviteit van notaties vergelijkend te evalueren (zie H2), (2) een gedetailleerde onderzoeksmethode voor de vergelijkende evaluatie van de intuïtiviteit van de PGA-modelleertaal (zie §3.2) en (3) een vergelijking van de evaluatieresultaten van de intuïtiviteit van de PGA-modelleertaal binnen de praktische bedrijfscontext (zie H4).

Er is gekozen voor een deductieve aanpak voor de onderzoeksmethode. Saunders et al. (2019, p. 801) geven hieraan de volgende omschrijving: “benadering van de theoretische aanpak waarbij een theoretisch voorstel wordt getest door de tewerkstelling van een onderzoeksstrategie die specifiek is ontworpen voor het testen ervan”. Met behulp van de deductieve aanpak wordt de theorie getest in het onderzoek. De deductieve aanpak bestaat uit de volgende vier stappen:

1. Start een theorie: het bestuderen van bestaande theorieën over conceptuele modellen (zie H2).
2. Stel een hypothese op: de onderzoeksvraag en deelvragen die als uitgangspunt dienen voor het onderzoek (zie §1.4) en de hypothesen (zie §3.2) om te onderzoeken of er een relatie bestaat tussen de afhankelijke en onafhankelijke variabelen.
3. Verzamel data tijdens het onderzoek: het verzamelen van evaluatieresultaten op basis van vragen m.b.t. geformuleerde hypothesen (zie H4).
4. Test de hypothesen: op basis van de verzamelde data tijdens het onderzoek (zie H4 en H5) worden de hypothesen bevestigd of ontkracht.

Als onderzoekstrategie is gekozen voor een experiment. Het doel van een experiment is het bestuderen van de mogelijkheid dat een onafhankelijke variabele een verandering veroorzaakt in afhankelijke variabelen. Saunders et al. (2019) beschrijft de onafhankelijke variabele als: “variabele die wordt gemanipuleerd of gewijzigd om de impact op een afhankelijke variabele te meten” (Saunders et al., 2019, p. 805) en de afhankelijke variabele als: “variabele die kan veranderen als gevolg van veranderingen in andere variabelen” (Saunders et al., 2019, p. 801). Een experiment maakt gebruik van hypothesen i.p.v. onderzoeksvragen, zodat de onderzoeker kan vaststellen of er een relatie bestaat tussen de variabelen. Saunders et al. (2019) beschrijven de nulhypothese als de verklaring dat er geen verschil of geen verband is tussen de variabelen en de alternatieve hypothese als de verklaring dat er wel een verschil of verband is tussen de variabelen.

3.2. Technisch ontwerp: uitwerking van de methode

Voor het evalueren van de intuïtiviteit van de PGA-notatie zijn drie afhankelijke variabelen geselecteerd. De variabelen voor het begrip van de PGA-notatie zijn vastgelegd in:

- interpretatieve effectiviteit: het argument om de interpretatieve effectiviteit (i.e. aantal correcte antwoorden op vragen) te meten komt voort uit het onderzoek van Burton-Jones et al. (2009), Gemino en Wand (2004), Recker et al. (2011), Cruzes et al. (2013) en Mendling et al. (2018). Het meten van de effectiviteit is belangrijk voor het verbeteren van het ontwerp van het conceptuele model.
- interpretatieve efficiëntie: het argument om de interpretatieve efficiëntie (i.e. tijd die nodig is om taken af te ronden) te meten komt voort uit het onderzoek van Gemino en Wand (2004), Shanks et al. (2008) en Mendling et al. (2018). Het meten van de efficiëntie is een belangrijke eigenschap van het conceptuele model.
- Intuïtieve begrijpelijkheid: het argument om de intuïtieve begrijpelijkheid (i.e. “het gemak waarmee gebruikers het diagram direct kunnen begrijpen is, zonder enige voorkennis of training”) te meten komt voor uit het onderzoek van Jošt et al. (2016, p. 91). Het meten van de representatievoorkeur (i.e. intuïtieve begrijpelijkheid) is een belangrijke toevoeging aan de experimentele methode voor het evalueren van de intuïtiviteit van de PGA-modelleertaal.

Onafhankelijke variabelen kunnen van invloed zijn op de afhankelijke variabelen. De onafhankelijke variabelen zijn:

- Originele PGA-notatie
- Nieuwe PGA-notatie

De metingen zijn:

1. Interpretatieve effectiviteit: het aantal correcte antwoorden op notatie-associatievragen, begripsvragen en herinneringsvragen.
2. Interpretatieve efficiëntie: de tijd (in seconden) die de participant nodig heeft om notatie-associatievragen, begripsvragen en herinneringsvragen af te ronden.
3. Representatievoorkeur: naar welke versie van de PGA-notatie gaat de intuïtieve voorkeur van de participant. De schaal loopt van 0 (e.g. ‘voorkeur gaat naar versie A’) tot 100 (e.g. ‘voorkeur gaat naar versie B’), waarbij 50 betekent dat er geen specifieke voorkeur is voor een versie.

De hypothesen zijn als volgt opgesteld om te onderzoeken of er een relatie bestaat tussen afhankelijke en onafhankelijke variabelen:

$H_{\text{effectiviteit}}$: de interpretatieve effectiviteit van de nieuwe PGA-notatie voor element x tijdens taak y is hoger dan die van de originele PGA-notatie voor element x tijdens taak y.

De verwachting is dat deze hypothese aanvaard kan worden en daarmee het aantal correcte antwoorden op vragen over de nieuwe PGA-notatie hoger is dan voor de originele PGA-notatie. Het experiment geeft zo een onafhankelijke waarneming over de effectiviteit van de nieuwe PGA-notatie.

$H_{\text{efficiëntie}}$: de interpretatieve efficiëntie van de nieuwe PGA-notatie voor element x tijdens taak Y is hoger dan die van de originele PGA-notatie voor element x tijdens taak y.

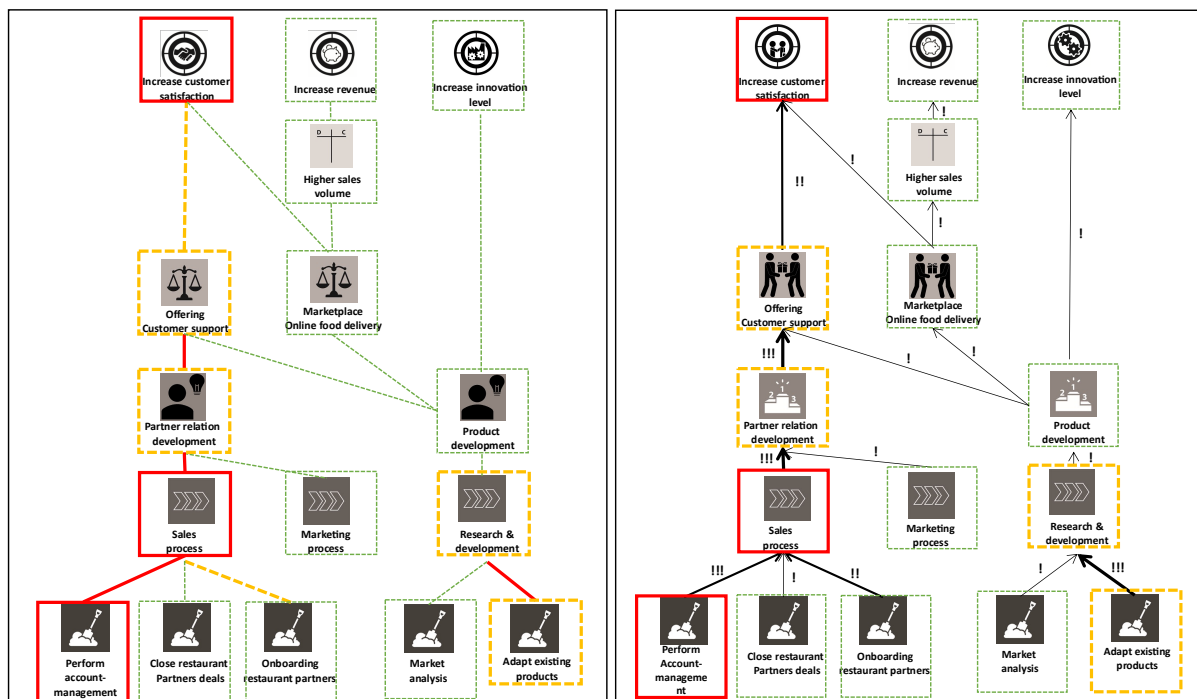
De verwachting is dat deze hypothese aanvaard kan worden en daarmee de tijd die de participant nodig heeft korter is om taken af te ronden voor de nieuwe PGA-notatie dan voor de originele PGA-notatie. Het experiment geeft zo een onafhankelijke waarneming over de efficiëntie van de nieuwe PGA-notatie.

$H_{representatievoorkeur}$: de representatievoorkeur van de nieuwe PGA-notatie voor element x is hoger dan die van de originele PGA-notatie voor element x tijdens taak y.

De verwachting is dat deze hypothese aanvaard is en daarmee de intuïtieve voorkeur wordt gegeven aan de nieuwe PGA-notatie dan voor de originele PGA-notatie. Het experiment geeft zo een onafhankelijke waarneming over de representatievoorkeur.

$H_{geaggregeerd}$: de geaggregeerde score van de nieuwe PGA-notatie voor element x is hoger dan die van de originele PGA-notatie voor element x. De verwachting is dat deze hypothese aanvaard is en daarmee de geaggregeerde score (i.e. de resultaten van de notatie-associatievragen, begripsvragen, herinneringsvragen en representatievoorkeursvragen geaggregeerd per PGA-element) hoger is voor de nieuwe PGA-notatie dan voor de originele PGA-notatie. De geaggregeerde score is zo een onafhankelijke waarneming over de intuïtieve voorkeur per element.

De casusorganisatie is een bedrijf dat een toonaangevende marktplaats biedt voor online voedsellevering. Het biedt klanten een gemakkelijke manier voor het bestellen en betalen van eten bij restaurantpartners, zoals schematisch in figuur 1 is weergegeven. Deze modellen worden vervolgens toegepast bij het beantwoorden van begripsvragen (zie figuur 2).













Figuur 1: Casus A (links) en casus B (rechts)

De gewenste participanten beoefenen verschillende functies, waaronder: manager, enterprise architect, solution architect, *business* analist en ontwerpers binnen een bedrijf en hebben affiniteit met IT en het strategisch belang van verschillende processen in te schatten.

Het onderzoek bestaat uit het vergelijken van vijf oorspronkelijke PGA-elementen en vijf nieuwe PGA-elementen om de intuïtieve van de PGA-moedeleertaal te evalueren en is een vervolgonderzoek op het artikel van Roelens en Bork (2020). De PGA-elementen die onderdeel zijn van het onderzoek zijn weergegeven in tabel 7.

Tabel 7: Oorspronkelijke & nieuwe PGA-elementen

PGA-element	Oorspronkelijke	Nieuw	Omschrijving
Internal goal			Strategic objective that describes a desired state or development of the company to ensure that a company excels in processes, decisions, and actions to meet its customers' expectations.
Value proposition			Offered set of products and/or services that provides value to the customers and other partners and competes in the overall value network.
Competence			An integrated and holistic set of knowledge, skills, and abilities, related to a specific set of resources, which is coordinated through processes to realize the intended value.
Value stream			Representation of the hierarchical structure, through which value is created at distinct levels in the business architecture. Value streams can be differentiated based on three levels of strategic importance: high, medium or low.
Customer goal			Strategic objective that describes a desired state or development of the company to ensure that a customer is satisfied with the products & services the company provides.

Het experiment bestaat uit het beantwoorden van contextgerichte vragen en vier taken rondom de PGA-notatie. Dit is schematisch weergegeven in figuur 2 en zijn beschreven in tabel 8.



Figuur 2: Taken in het experiment

Tabel 8: Survey taken in het onderzoek

Taak	Omschrijving
Notatie-associatievragen	De PGA-notatie wordt getoond met als doel om tot drie intuïtieve associaties te komen, per element als ze worden bekeken.
Begripsvragen	Het beantwoorden van een reeks begripsvragen over PGA-elementen in het conceptuele model.
Herinneringsvragen	De PGA-notatie en de onderliggende betekenis wordt getoond en vervolgens weer weggelaten. Aan de participanten wordt gevraagd om daarna de visualisatie te koppelen aan het juiste element in het model.
Representatievoorkeurvragen	De originele en nieuwe PGA-notatie elementen worden getoond. Aan de participanten wordt gevraagd waar de intuïtieve voorkeur naar uitgaat.

Naast het vastleggen van de resultaten van de vragen worden voor de taken notatie-associatievragen, begripsvragen en herinneringsvragen ook automatisch de tijden bij gehouden hoe lang een participant nodig heeft om de taak uit te voeren per element.

In tabel 9 is vervolgens de relatie weergegeven van de afhankelijke variabelen en de uitgevoerde taken in het onderzoek.

Tabel 9: Relatie afhankelijke variabelen tot de toets

Afhankelijke variabelen	Taak	Referentie	Toelichting
Interpretatieve effectiviteit	Notatie-associatievragen, begripsvragen, herinneringsvragen,	Gemino en Wand (2004), Recker et al. (2011), Burton-Jones et al. (2009), Cruzes et al. (2013) en Mendling et al. (2018)	De taken notatie-associatievragen, begripsvragen, herinneringsvragen zijn geschikt voor het meten van de interpretatieve effectiviteit (i.e. aantal correcte antwoorden op vragen). De resultaten geven inzicht in welke PGA-element de meeste aantal correcte antwoorden krijgt.
Interpretatieve efficiëntie	Notatie-associatievragen, begripsvragen, herinneringsvragen (tijden)	Gemino en Wand (2004), Shanks et al. (2008) en Mendling et al. (2018)	De taken notatie-associatievragen, begripsvragen, herinneringsvragen zijn geschikt voor het meten van de interpretatieve efficiëntie (i.e. tijd die nodig is om taken af te ronden). Deze resultaten geven inzicht in aan welk PGA-element minder tijd wordt besteed bij het formuleren van een antwoord en daarmee efficiëntie behaald kan worden.
Representatievoorkeur	Representatievoorkeur	Jošt et al. (2016, p. 91).	De taak representatievoorkeur is geschikt voor het meten van de intuïtieve begrijpelijkheid (i.e. het gemak waarmee gebruikers het diagram direct kunnen begrijpen, zonder enige voorkennis of training). Deze taak draagt bij aan het evalueren van de intuïtiviteit van het PGA-element.

Voor het onderzoek is gekozen voor mix van *between-subjects design en within-subjects design*. *Between-subjects*: Concreet betekent dit dat er gebruikt gemaakt wordt van twee subgroepen, waarin de notaties worden gewisseld. De participanten worden op basis van het geboortjaar (i.e. even of oneven) aan één van de twee subgroepen toegewezen. Beide groepen krijgen dezelfde taken, maar er wordt een andere combinatie van de oorspronkelijke en nieuwe PGA-elementen gebruikt. Dit is toegepast voor de notatie-associatievragen, begripsvragen en de herinneringsvragen. Hiervoor is gekozen om de participanten niet op voorhand inzicht te geven in welke PGA-notatie de oorspronkelijke of nieuwe is.

Voor de representatievoorkeursvragen is gekozen voor *Within-subjects*. Hiervoor is gekozen om de kracht van de experimentele test, ondanks de beperkt aantal participanten, te maximaliseren (Saunders et al., 2019).

De experimentele taken binnen het experiment zijn dus als volgt:

- Participanten beantwoorden context vragen (Mendling et al., 2018)
 - *Gender: What is your gender?*
 - *Year of Birth: Please select your year of birth*
 - *Highest degree: What is your highest degree?*
 - *Job sector: In which sector are you working?*
 - *Modeling experience: Which conceptual modeling languages do you know?*
 - *Modeling expertise: Please rate your level of modeling expertise.*
 - *Model number read: How many conceptual models have you read by average during the last five years?*
 - *Model number made: How many conceptual models have you made by average during the last five years?*
- Participanten beantwoorden notatie-associatievragen ((Roelens et al., 2019), (Roelens & Bork, 2020) en (Bork et al., 2019))
 - De PGA-notatie wordt getoond met als doel om tot drie intuïtieve associaties te komen wanneer ze worden bekeken. Belangrijk is dat deze notaties gepresenteerd worden zonder enige hint van bijvoorbeeld de naam of de betekenis van het concept.
- Participanten beantwoorden begripsvragen op basis van een PGA-model over een strategisch proces ((Gemino & Wand, 2004), (Recker et al., 2011), (Burton-Jones et al., 2009), (Cruzes et al., 2013) en (Mendling et al., 2018))

- Het beantwoorden van een reeks begripsvragen over PGA-elementen in het conceptuele model die in overweging genomen worden door de participanten.
- Participanten beantwoorden herinneringsvragen (Gemino & Wand, 2004)
 - De PGA-notatie en onderliggende betekenis wordt getoond en vervolgens weer weggelaten. De participanten wordt gevraagd om daarna de visualisatie te koppelen aan het juiste element in het model.
- Participanten beantwoorden representatievoorkeursvragen (Jošt et al., 2016)
 - De originele en nieuwe PGA-notatie elementen worden getoond. Aan de participanten wordt gevraagd waar de intuïtieve voorkeur naar uit gaat.

De procedure voor het uitvoeren van het experiment is dat het digitaal wordt afgenomen. Hiervoor is gekozen om de meting interpretatieve efficiëntie goed te kunnen meten, door het automatisch bijhouden van de tijd die nodig is voor het beantwoorden van een vraag. Voordat het experiment kan starten wordt eerst een pretest uitgevoerd met twee participanten die geen onderdeel uitmaken van het experiment. Vervolgens wordt het experiment uitgevoerd en worden participanten op basis van het geboortjaar (i.e. even of oneven) aan één van de twee groepen toegewezen. Daarna wordt de gegevensverzameling beoordeeld en eventuele participanten worden uitgesloten als blijkt dat deze maar een gedeelte van het experiment hebben uitgevoerd.

3.3. Gegevensanalyse

Om antwoord te kunnen geven op de hoofdvraag en de drie deelvragen is het nodig om de eigen gegevens te analyseren. Voor het empirisch analyseren van de gegevens zijn er verschillende statistische toetsen. Het onderzoek gaat uit van het empirisch evalueren van de PGA-methode en hiervoor wordt gebruik gemaakt van ongepaarde groepen m.u.v. de representatievoorkeursvragen. Voor de representatievoorkeursvragen wordt gebruikt gemaakt van één homogene groep, waarbij iedereen dezelfde vragen en PGA-notaties gekregen.

Tabel 10 beschrijft de relatie tussen de afhankelijke variabelen en de toegepaste statistische toetsen.

Tabel 10: Relatie afhankelijke variabelen tot de toetsen

Afhankelijke variabelen	Taak	Groep	Type data	Toetsen
Interpretatieve effectiviteit	Notatie-associatievragen, begripsvragen, herinneringsvragen,	Ongepaard	1. Normaal verdeeld 2. Niet normaal verdeeld/ordinaal	1. Independent Samples Test 2. Mann-Whitney U Test
Interpretatieve efficiëntie	Notatie-associatievragen, begripsvragen, herinneringsvragen (tijden)	Ongepaard	1. Normaal verdeeld 2. Niet normaal verdeeld/ordinaal	1. Independent Samples Test 2. Mann-Whitney U Test
Representatievoorkeur	Representatievoorkeur	Homogene groep	1. Normaal verdeeld 2. Niet normaal verdeeld	1. One-Sample Test 2. One-Sample Wilcoxon Signed Rank Test

Bij het toetsen van de interpretatieve effectiviteit en interpretatieve efficiëntie wordt gebruikt gemaakt van ongepaarde groepen. Als de data normaal verdeeld is dan wordt de ongepaarde t-toets gebruikt om de gemiddelden van twee onafhankelijke groepen met elkaar te vergelijken (Versnellingsplan, 2020). Wanneer de data niet normaal verdeeld is dan wordt *de Mann-Whitney U*

toets gebruikt om te toetsen of de gemiddelde rangnummers van de verdelingen van twee ongepaarde groepen van elkaar verschillen (Versnellingsplan, 2020).

Bij het toetsen van de representatievoorkeur wordt gebruikt gemaakt van een homogene groep. Als de data normaal verdeeld is dan wordt de *One-Sample Test* gebruikt om het gemiddelde van de steekproef te vergelijken met een bekend gemiddelde of norm van de totale populatie (Versnellingsplan, 2020). Wanneer de data niet normaal verdeeld is dan wordt de *One-Sample Wilcoxon Signed Rank Test* toets gebruikt om de mediaan van een steekproef te vergelijken met een bekende mediaan (Versnellingsplan, 2020).

3.4. Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten

Dit hoofdstuk beschrijft de reflectie t.a.v. de validiteit, betrouwbaarheid en de ethische aspecten. De beperkingen met betrekking tot de validiteit, betrouwbaarheid en de ethische aspecten worden in hoofdstuk 5 geëvalueerd.

Construct validiteit

Het onderzoek betreft een kwantitatief onderzoek en geeft antwoorden op hoe twee verschillende versies van de PGA-notaties geëvalueerd kunnen worden. Om de construct validiteit te waarborgen voor het empirisch analyseren van de PGA-notatie is gebruik gemaakt van drie afhankelijke variabelen uiteengezet in vier taken in de survey. Deze taken zijn zo gekozen dat ze de afhankelijke variabelen juist meten (zie §3.2.). De verkregen informatie wordt daarna gecontroleerd of het survey volledig en correct is ingevuld, omdat er meetfouten in kunnen zitten. Meetfouten kunnen opgelost worden door bijvoorbeeld een participant uit te sluiten of het gehele experiment opnieuw te doen.

Interne validiteit

Er zullen ca. 15 experimenten uitgevoerd worden om de intuïtiviteit van de originele PGA-notatie en de nieuwe PGA-notatie te evalueren binnen een praktische bedrijfscontext. Naast de persoonlijke resultaten worden ook de data van twee andere studenten gebruikt. Dit betekent dat er twee bronnen zijn, namelijk een eigen dataset van ca. 15 participanten en een gecombineerde dataset van ca. 45 participanten. Hierbij wordt gebruik gemaakt van de breed geaccepteerde applicatie SPSS voor het uitvoeren van statistische toetsen. De hypothesen worden statistisch getoetst. Hierbij moet een kanttekening geplaatst worden dat het aantal participanten beperkt is in dit onderzoek. Om verstorende factoren uit te sluiten krijgen beide groepen dezelfde taken, maar wordt een andere combinatie gebruikt van de oorspronkelijke en nieuwe PGA-elementen m.u.v. de representatievoorkeursvragen. Voor de representatievoorkeursvragen krijgen de participanten dezelfde vragen en PGA-notaties. Door de beschreven procedure te volgen, zoals weergegeven in §3.2, wordt het mogelijke leereffect zoveel mogelijk voorkomen. Hiervoor hebben participanten niet op voorhand inzicht welke PGA-notatie de oorspronkelijke of nieuwe is.

Externe validiteit

De resultaten van het onderzoek zijn gebaseerd op de strategische bedrijfsprocessen van een bedrijf dat een toonaangevende marktplaats biedt voor online voedsellevering. Dit model wordt toegepast bij begripvragen. De participanten hebben verschillende functies, waaronder manager, enterprise architect, solution architect, *business* analist en ontwerpers binnen een bedrijf en hebben affiniteit met IT en strategische bedrijfsprocessen. Hiervoor is gekozen, omdat de participanten allemaal wel bekend zijn met de processen rondom een marktplaats van restaurants om eten te kunnen bestellen. Met deze achtergrondkennis zijn de participanten in staat om het strategisch belang van de verschillende strategische bedrijfsprocessen in te kunnen schatten, zonder dat daarvoor inhoudelijk kennis nodig is. De resultaten uit dit onderzoek kunnen worden gegeneraliseerd op het gebied van de intuïtiviteit van de PGA-notatie door het in kaart brengen van strategische bedrijfsprocessen in een logistieke keten.

Betrouwbaarheid

Triangulatie is “het samenvoegen van twee of meer onafhankelijke databronnen of dataverzamelmethode binnen één onderzoek om ervoor te zorgen dat de weergegeven gegevens vertellen wat het onderzoek weergeeft” (Saunders et al., 2019, p. 819). In het onderzoek wordt data verzameld van individuele participanten, die vervolgens zijn ingedeeld in groep A of groep B. Daarnaast wordt er bij representatievoorkeursvragen ook een toelichting gevraagd en levert een extra dataverzameling op. De indeling is op basis van een beschreven en repliceerbare onderzoeksmethode (zie §3.2). Vanuit verschillende invalshoeken is onderzoek gedaan om de onderzoeksvraag te kunnen beantwoorden door gebruik te maken van zowel onafhankelijke variabelen (i.e. oorspronkelijke PGA-notatie, nieuwe PGA-notatie) als afhankelijke variabelen (i.e. interpretatieve effectiviteit, interpretatieve efficiëntie en intuïtieve begrijpelijkheid).

Ethische aspecten

Tijdens het onderzoek wordt gewerkt met een *Informed Consent*. Hierin wordt vastgelegd dat de onderzoeker zich verplicht tot het vertrouwelijk behandelen van data en dat deze volledig geanonimiseerd wordt. Ook mogen participanten op elk moment stoppen met deelname aan het experiment. Het *Informed Consent* zal digitaal versterkt worden als onderdeel van het experiment. Verwerking en opslag van persoonsgegevens zal plaatsvinden in overeenstemming met de algemene verordening gegevensbescherming (i.e. AVG). Voor de werving van de participanten is gebruikt gemaakt van het eigen netwerk met als doelgroep IT managers, enterprise architecten, solution architecten, business analisten en ontwerpers. Er is geen beloning voor deelname aan het experiment en deelname is op vrijwillige basis. Het experiment zal digitaal worden afgenomen door gebruik te maken van een enquête-tool. De resultaten kunnen hierin geraadpleegd worden. Al het experimenteel materiaal zal ter verificatie worden ingediend bij de Open Universiteit.

4. Resultaten

Dit hoofdstuk beschrijft de uitvoering van het onderzoek en de resultaten om antwoord te kunnen geven of de nieuwe versie van de PGA-notatie intuïtiever is dan de oorspronkelijke versie (i.e. deelonderzoeksvraag 3).

4.1. Uitvoering van het onderzoek

Bijlage D beschrijft een overzicht van de bijlagen die onderdeel uitmaken van de resultaten. Het betreft de export van de survey en de dataset voor SPSS. Ook beschrijft bijlage B een lijst van elementen die gekoppeld zijn aan het label dat gebruikt is in SPSS om de vraag te identificeren. Hierdoor is het mogelijk om de uitkomsten te linken aan de originele vraag.

Tijdens de uitvoering van het onderzoek is gewerkt met een dubbele analyse. Het experiment is opgedeeld in persoonlijke resultaten en groepsresultaten (i.e. persoonlijke resultaten gecombineerd met de uitkomsten van twee andere studenten).

Persoonlijke resultaten

Aan het experiment hebben in totaal 14 participanten meegedaan die werkzaam zijn in de particuliere sector (i.e. 43%) of publieke sector (i.e. 57%). 86% van de respondenten heeft een bachelor of hoger niveau. De respondenten zijn niet bekend met de PGA-methode. In totaal heeft 64% enige kennis van modellering (i.e. beoordeeld als gemiddeld, hoog of zeer hoog). Het aantal respondenten die meer dan 20 modellen hebben gelezen in de afgelopen vijf jaar is 57% en 36% heeft in de afgelopen vijf jaar meer dan 20 modellen gemaakt.

Groepsresultaten

Aan het experiment hebben in totaal 44 participanten meegedaan die werkzaam zijn in de particuliere sector (i.e. 68%), publieke sector (i.e. 27%), wetenschappelijke sector (i.e. 2%) of anders (i.e. 2%). 95% van de respondenten heeft een bachelor of hoger niveau. De respondenten zijn niet bekend met de PGA-methode. In totaal heeft 59% enige kennis van modellering (i.e. beoordeeld als gemiddeld, hoog of zeer hoog). Het aantal respondenten die meer dan 20 modellen hebben gelezen in de afgelopen vijf jaar is 45% en 23% heeft in de afgelopen vijf jaar meer dan 20 modellen gemaakt.

4.2. Uitkomsten: effectiviteit

Voor de taken notatie-associatievragen, begripsvragen en herinneringsvragen zijn het aantal correcte antwoorden gemeten op vragen om de interpretatieve effectiviteit (zie §3.2) te meten.

In tabel 11 worden de persoonlijke resultaten en in tabel 12 worden de groepsresultaten weergegeven. De gevolgde testprocedure staat in bijlage O beschreven.

De bijbehorende hypothesen zijn als volgt:

- H_0 : de interpretatieve effectiviteit van de nieuwe PGA-notatie voor element x tijdens taak y is gelijk aan die van de originele PGA-notatie voor element x tijdens taak y .
- $H_{\text{effectiviteit}}$: de interpretatieve effectiviteit van de nieuwe PGA-notatie voor element x tijdens taak y is hoger dan die van de originele PGA-notatie voor element x tijdens taak y .

Als de toets aangeeft dat er een significant verschil is (e.g. er is meer aan de hand dan toeval) dan wordt de alternatieve hypothese ($H_{\text{effectiviteit}}$) aanvaard en de nulhypothese (H_0) wordt verworpen. Een significant verschil is als de p -value ≤ 0.05 is. Wanneer de toets geen significant verschil aangeeft dan wordt de alternatieve hypothese ($H_{\text{effectiviteit}}$) niet aanvaard en wordt de nulhypothese (H_0) niet verworpen.

De uitkomsten van de persoonlijke resultaten in tabel 11 voor de taken notatie-associatievragen, begripsvragen en herinneringsvragen laten zien dat er geen significante verschillen zijn voor de afhankelijke variabelen *internal goal*, *value proposition*, *competence*, *value stream* en *customer goal*.

Tabel 11: Persoonlijke resultaten: uitkomst meting effectiviteit

Taak (y)	Afhankelijke variabele (x)	Label survey	Toets	Mean Rank (0/1)	Asymp. Sig. (2-tailed)	Sig. (1-tailed)	H ₀	H _{effectiviteit}	Bron
Notatie-associatievragen	<i>Internal goal</i>	AV: Notation Association - Q1	Independent Samples Test	.143/-.048	.486	.243	Niet verwerpen	Niet aanvaarden	Bijlage E: H3
Notatie-associatievragen	<i>Value proposition</i>	AV: Notation Association - Q2	Mann-Whitney U Test	6.50/8.50	.141	.071	Niet verwerpen	Niet aanvaarden	Bijlage E: H2
Notatie-associatievragen	<i>Competence</i>	AV: Notation Association - Q3	Mann-Whitney U Test	8.07/6.93	.476	.238	Niet verwerpen	Niet aanvaarden	Bijlage E: H2
Notatie-associatievragen	<i>Value stream</i>	AV: Notation Association - Q4	Mann-Whitney U Test	8.43/6.75	.332	.834	Niet verwerpen	Niet aanvaarden	Bijlage E: H2
Notatie-associatievragen	<i>Customer goal</i>	AV: Notation Association - Q5	Independent Samples Test	.405/.405	1.000	.500	Niet verwerpen	Niet aanvaarden	Bijlage E: H3
Begripsvragen	<i>Customer goal</i>	CQQ1Code	Mann-Whitney U Test	8.00/7.00	.317	.842	Niet verwerpen	Niet aanvaarden	Bijlage G: H2
Begripsvragen	<i>Value proposition</i>	CQQ2Code	Mann-Whitney U Test	8.00/7.00	.606	.697	Niet verwerpen	Niet aanvaarden	Bijlage G: H2
Begripsvragen	<i>Value stream</i>	CQQ3Code	Mann-Whitney U Test	7.50/7.50	1.000	.500	Niet verwerpen	Niet aanvaarden	Bijlage G: H2
Begripsvragen	<i>Competence</i>	CQQ4Code	Mann-Whitney U Test	8.00/7.00	.591	.296	Niet verwerpen	Niet aanvaarden	Bijlage G: H2
Begripsvragen	<i>Internal goal</i>	CQQ5Code	Mann-Whitney U Test	7.00/8.00	.317	.842	Niet verwerpen	Niet aanvaarden	Bijlage G: H2
Begripsvragen	<i>Value stream</i>	CQQ6Code	Mann-Whitney U Test	8.00/7.00	.317	.842	Niet verwerpen	Niet aanvaarden	Bijlage G: H2
Herinnerings-Vragen	<i>Internal goal</i>	RQQ1Code	Mann-Whitney U Test	6.57/8.43	.297	.852	Niet verwerpen	Niet aanvaarden	Bijlage I: H2
Herinnerings-Vragen	<i>Value proposition</i>	RQQ2Code	Mann-Whitney U Test	9.64/5.36	.033	.835	Niet verwerpen	Niet aanvaarden	Bijlage I: H2
Herinnerings-Vragen	<i>Competence</i>	RQQ3Code	Mann-Whitney U Test	8.14/6.86	.496	.248	Niet verwerpen	Niet aanvaarden	Bijlage I: H2
Herinnerings-Vragen	<i>Value stream</i>	RQQ4Code	Mann-Whitney U Test	8.07/6.93	.476	.762	Niet verwerpen	Niet aanvaarden	Bijlage I: H2
Herinnerings-Vragen	<i>Customer goal</i>	RQQ5Code	Mann-Whitney U Test	9.07/5.93	.114	.943	Niet verwerpen	Niet aanvaarden	Bijlage I: H2

De uitkomsten van de groepsresultaten in tabel 12 voor de taken begripsvragen en herinneringsvragen laten zien dat er geen significante verschillen zijn voor de afhankelijke variabelen *internal goal*, *value proposition*, *competence*, *value stream* en *customer goal*. Voor de taak notatie-associatievragen is er een significant verschil voor de afhankelijke variabele *value proposition*. Voor de afhankelijk variabelen *internal goal*, *competence*, *value stream*, *customer goal* is er geen significant verschil.

Tabel 12: Groepsresultaten: uitkomst meting effectiviteit

Taak (y)	Afhankelijke variabele (x)	Label survey	Toets	Mean Rank (0/1)	Asymp. Sig. (2-tailed)	Sig. (1-tailed)	H ₀	H _{effectiviteit}	Bron
Notatie-associatievragen	<i>Internal goal</i>	AV: Notation Association - Q1	Mann-Whitney U Test	23.13/21.75	.668	.334	Niet verwerpen	Niet aanvaarden	Bijlage F: H2
Notatie-associatievragen	<i>Value proposition</i>	AV: Notation Association – Q2	Mann-Whitney U Test	21.00/24.30	.052	.026	Verwerpen	Aanvaarden	Bijlage F: H2
Notatie-associatievragen	<i>Competence</i>	AV: Notation Association – Q3	Mann-Whitney U Test	22.85/22.08	.647	.324	Niet verwerpen	Niet aanvaarden	Bijlage F: H2
Notatie-associatievragen	<i>Value Stream</i>	AV: Notation Association – Q4	Mann-Whitney U Test	24.23/20.43	.225	.888	Niet verwerpen	Niet aanvaarden	Bijlage F: H2
Notatie-associatievragen	<i>Customer goal</i>	AV: Notation Association – Q5	Mann-Whitney U Test	23.10/21.78	.684	.658	Niet verwerpen	Niet aanvaarden	Bijlage F: H2
Begripsvragen	<i>Customer goal</i>	CQQ1Code	Mann-Whitney U Test	23.50/21.30	.117	.942	Niet aanvaarden	Niet aanvaarden	Bijlage H: H2
Begripsvragen	<i>Value proposition</i>	CQQ2Code	Mann-Whitney U Test	26.25/18.00	.014	.993	Niet aanvaarden	Niet aanvaarden	Bijlage H: H2
Begripsvragen	<i>Value stream</i>	CQQ3Code	Mann-Whitney U Test	22.58/22.40	.896	.552	Niet aanvaarden	Niet aanvaarden	Bijlage H: H2
Begripsvragen	<i>Competence</i>	CQQ4Code	Mann-Whitney U Test	22.33/22.70	.912	.544	Niet aanvaarden	Niet aanvaarden	Bijlage H: H2
Begripsvragen	<i>Internal goal</i>	CQQ5Code	Mann-Whitney U Test	21.42/23.80	.333	.835	Niet aanvaarden	Niet aanvaarden	Bijlage H: H2
Begripsvragen	<i>Value stream</i>	CQQ6Code	Mann-Whitney U Test	22.50/22.50	1.000	.500	Niet aanvaarden	Niet aanvaarden	Bijlage H: H2
Herinnerings-Vragen	<i>Internal goal</i>	RQQ1Code	Mann-Whitney U Test	22.17/22.90	.838	.581	Niet verwerpen	Niet aanvaarden	Bijlage J: H2
Herinnerings-Vragen	<i>Value proposition</i>	RQQ2Code	Mann-Whitney U Test	25.67/18.70	.051	.975	Niet verwerpen	Niet aanvaarden	Bijlage J: H2
Herinnerings-Vragen	<i>Competence</i>	RQQ3Code	Mann-Whitney U Test	23.63/21.15	.466	.233	Niet verwerpen	Niet aanvaarden	Bijlage J: H2
Herinnerings-Vragen	<i>Value steam</i>	RQQ4Code	Mann-Whitney U Test	25.58/18.80	.042	.979	Niet verwerpen	Niet aanvaarden	Bijlage J: H2
Herinnerings-Vragen	<i>Customer goal</i>	RQQ5Code	Mann-Whitney U Test	26.19/18.08	.025	.988	Niet verwerpen	Niet aanvaarden	Bijlage J: H2

4.3. Uitkomsten: efficiëntie

Voor de taken notatie-associatievragen, begripsvragen en herinneringsvragen is de tijd gemeten (in seconden) die de participant nodig heeft om taken af te ronden om zo de interpretatieve efficiëntie (zie §3.2) te meten. Hoe minder tijd de participant nodig heeft hoe efficiënter het resultaat is.

In tabel 13 worden de resultaten weergegeven voor de persoonlijke resultaten en in tabel 14 worden de groepsresultaten weergegeven. De gevolgde testprocedure staat in bijlage O beschreven.

De bijbehorende hypothesen zijn als volgt:

- H₀: de interpretatieve effectiviteit van de nieuwe PGA-notatie voor element x tijdens taak y is gelijk aan die van de originele PGA-notatie voor element x tijdens taak y.

- $H_{\text{efficiëntie}}$: de interpretatieve efficiëntie van de nieuwe PGA-notatie voor element x tijdens taak Y is hoger dan die van de originele PGA-notatie voor element x tijdens taak y.

Als de toets aangeeft dat er een significant verschil is (e.g. er is meer aan de hand dan toeval) dan wordt de alternatieve hypothese ($H_{\text{efficiëntie}}$) aanvaard en de nulhypothese (H_0) wordt verworpen. Een significant verschil is als de p-value ≤ 0.05 is. Wanneer de toets geen significant verschil aangeeft dan wordt de alternatieve hypothese ($H_{\text{efficiëntie}}$) niet aanvaard en blijft de nulhypothese (H_0) niet verworpen.

De uitkomsten van de persoonlijke resultaten in tabel 13 voor de taken notatie-associatievragen, begripsvragen en herinneringsvragen laten zien dat er geen significante verschillen zijn voor de afhankelijke variabelen *internal goal*, *value proposition*, *competence*, *value stream* en *customer goal*.

Tabel 13: Persoonlijke resultaten: uitkomst meting efficiëntie tijden

Taak (y)	Afhankelijke variabele (x)	Label survey	Toets	Mean Rank (0/1)	Asymp. Sig. (2-tailed)	Sig. (1-tailed)	H_0	$H_{\text{efficiëntie}}$	Bron
Notatie-associatievragen	<i>Internal goal</i>	Group time: Notation association – Q1	<i>Independent Samples Test</i>	94.37/68.08	.332	.834	Niet verworpen	Niet aanvaarden	Bijlage E
Notatie-associatievragen	<i>Value proposition</i>	Group time: Notation association - Q2	Mann-Whitney U Test	7.00/8.00	.655	.673	Niet verworpen	Niet aanvaarden	Bijlage E
Notatie-associatievragen	<i>Competence</i>	Group time: Notation association - Q3	Mann-Whitney U Test	7.14/7.86	.749	.375	Niet verworpen	Niet aanvaarden	Bijlage E
Notatie-associatievragen	<i>Value Stream</i>	Group time: Notation association - Q4	Mann-Whitney U Test	8.29/6.71	.482	.241	Niet verworpen	Niet aanvaarden	Bijlage E
Notatie-associatievragen	<i>Customer goal</i>	Group time: Notation association - Q5	Mann-Whitney U Test	6.57/8.43	.406	.797	Niet verworpen	Niet aanvaarden	Bijlage E
Begripsvragen	<i>Customer goal</i>	Group time: Comprehension questions – Q1	<i>Independent Samples Test</i>	46.36/42.36	.720	.360	Niet verworpen	Niet aanvaarden	Bijlage G
Begripsvragen	<i>Value proposition</i>	Group time: Comprehension questions - Q2	Mann-Whitney U Test	7.57/7.43	.949	.475	Niet verworpen	Niet aanvaarden	Bijlage G
Begripsvragen	<i>Value stream</i>	Group time: Comprehension questions - Q3	Mann-Whitney U Test	6.57/8.43	.406	.797	Niet verworpen	Niet aanvaarden	Bijlage G
Begripsvragen	<i>Competence</i>	Group time: Comprehension questions – Q4	<i>Independent Samples Test</i>	93.20/92.11	.976	.512	Niet verworpen	Niet aanvaarden	Bijlage G
Begripsvragen	<i>Internal goal</i>	Group time: Comprehension questions - Q5	Mann-Whitney U Test	6.00/9.00	.180	0.09	Niet verworpen	Niet aanvaarden	Bijlage G
Begripsvragen	<i>Value Stream</i>	Group time: Comprehension questions – Q6	<i>Independent Samples Test</i>	39.07/39.19	.987	.507	Niet verworpen	Niet aanvaarden	Bijlage G
Herinnerings-Vragen	<i>Internal goal</i>	Group time: Recall questions – Q1	<i>Independent Samples Test</i>	17.00/16.80	.956	.522	Niet verworpen	Niet aanvaarden	Bijlage I
Herinnerings-Vragen	<i>Value proposition</i>	Group time: Recall questions – Q2	<i>Independent Samples Test</i>	15.67/14.73	.815	.408	Niet verworpen	Niet aanvaarden	Bijlage I
Herinnerings-Vragen	<i>Competence</i>	Group time: Recall questions – Q3	Mann-Whitney U Test	6.14/8.86	.225	.113	Niet verworpen	Niet aanvaarden	Bijlage I
Herinnerings-Vragen	<i>Value stream</i>	Group time: Recall questions – Q4	Mann-Whitney U Test	7.00/8.00	.655	.673	Niet verworpen	Niet aanvaarden	Bijlage I

Taak (y)	Afhankelijke variabele (x)	Label survey	Toets	Mean Rank (0/1)	Asymp. Sig. (2-tailed)	Sig. (1-tailed)	H ₀	H _{efficiëntie}	Bron
Herinnerings-Vragen	<i>Customer goal</i>	Group time: Recall questions – Q5	Independent Samples Test	11.02/14.80	.209	.896	Niet verwerpen	Niet aanvaarden	Bijlage I

De uitkomsten van de groepsresultaten in tabel 14 voor de taak notatie-associatievragen laten zien dat er geen significante verschillen zijn voor de afhankelijke variabelen *internal goal*, *value proposition*, *competence*, *value stream* en *customer goal*. Voor de taak begripsvragen is er een significant verschil voor de afhankelijke variabele *value stream* (i.e. *Group time: Comprehension questions – Q6*). Voor de afhankelijk variabelen *internal goal*, *value proposition*, *competence*, *value stream* (i.e. *Group time: Comprehension questions – Q3*), *customer goal* is er geen significant verschil voor de taak begripsvragen. Voor de taak herinneringsvragen is er een significant verschil voor de afhankelijke variabele *internal goal*, *competence* en voor de afhankelijk variabelen *value proposition*, *value stream*, *customer goal* is er geen significant verschil.

Tabel 14: Groepsresultaten: uitkomst meting efficiëntie tijden

Taak (y)	Afhankelijke variabele (x)	Label survey	Toets	Mean Rank (0/1)	Asymp. Sig. (2-tailed)	Sig. (1-tailed)	H ₀	H _{efficiëntie}	Bron
Notatie-associatievragen	<i>Internal goal</i>	Group time: Notation association – Q1	Mann-Whitney U Test	21.92/23.20	.741	.371	Niet verwerpen	Niet aanvaarden	Bijlage F
Notatie-associatievragen	<i>Value proposition</i>	Group time: Notation association – Q2	Mann-Whitney U Test	20.58/24.80	.278	.861	Niet verwerpen	Niet aanvaarden	Bijlage F
Notatie-associatievragen	<i>Competence</i>	Group time: Notation association – Q3	Mann-Whitney U Test	22.75/22.20	.888	.556	Niet verwerpen	Niet aanvaarden	Bijlage F
Notatie-associatievragen	<i>Value Stream</i>	Group time: Notation association – Q4	Mann-Whitney U Test	24.88/19.65	.888	.444	Niet verwerpen	Niet aanvaarden	Bijlage F
Notatie-associatievragen	<i>Customer goal</i>	Group time: Notation association – Q5	Mann-Whitney U Test	20.17/25.30	.187	.907	Niet verwerpen	Niet aanvaarden	Bijlage F
Begripsvragen	<i>Customer goal</i>	Group time: Comprehension questions - Q1	Mann-Whitney U Test	22.13/22.95	.832	.584	Niet verwerpen	Niet aanvaarden	Bijlage H
Begripsvragen	<i>Value proposition</i>	Group time: Comprehension questions – Q2	Mann-Whitney U Test	22.67/22.30	.925	.463	Niet verwerpen	Niet aanvaarden	Bijlage H
Begripsvragen	<i>Value stream</i>	Group time: Comprehension questions – Q3	Mann-Whitney U Test	20.69/24.68	.305	.848	Niet verwerpen	Niet aanvaarden	Bijlage H
Begripsvragen	<i>Competence</i>	Group time: Comprehension questions – Q4	Mann-Whitney U Test	20.00/25.50	.157	.079	Niet verwerpen	Niet aanvaarden	Bijlage H
Begripsvragen	<i>Internal goal</i>	Group time: Comprehension questions – Q5	Mann-Whitney U Test	21.13/24.15	.437	.219	Niet verwerpen	Niet aanvaarden	Bijlage H
Begripsvragen	<i>Value Stream</i>	Group time: Comprehension questions – Q6	Mann-Whitney U Test	27.17/16.90	.008	.004	Verwerpen	Aanvaarden	Bijlage H
Herinnerings-vragen	<i>Internal goal</i>	Group time: Recall questions - Q1	Mann-Whitney U Test	19.50/26.10	.090	.045	Verwerpen	Aanvaarden	Bijlage J
Herinnerings-vragen	<i>Value proposition</i>	Group time: Recall questions - Q2	Mann-Whitney U Test	23.71/21.05	.494	.247	Niet verwerpen	Niet aanvaarden	Bijlage J

Taak (y)	Afhankelijke variabele (x)	Label survey	Toets	Mean Rank (0/1)	Asymp. Sig. (2-tailed)	Sig. (1-tailed)	H ₀	H _{efficiëntie}	Bron
Herinneringsvragen	Competence	Group time: Recall questions - Q3	Mann-Whitney U Test	19.58/26.00	.099	.0495	Verwerpen	Aanvaarden	Bijlage J
Herinneringsvragen	Value stream	Group time: Recall questions - Q4	Mann-Whitney U Test	21.77/23.38	.680	.660	Niet verwerpen	Niet aanvaarden	Bijlage J
Herinneringsvragen	Customer goal	Group time: Recall questions - Q5	Mann-Whitney U Test	21.02/24.28	.403	.799	Niet verwerpen	Niet aanvaarden	Bijlage J

4.4. Uitkomsten: representatievoorkeur

Voor de representatievoorkeurstaak is de intuïtieve voorkeur van de respondenten gemeten voor de oorspronkelijke notatie of de nieuwe notatie (zie §3.2). Als de gemiddelde/mediaan score groter is dan 50, dan is er een voorkeur voor de nieuwe notatie.

Aan de participanten is een extra mogelijkheid geboden om een optionele toelichting te geven per PGA-element welke is opgenomen in bijlage L. De toelichtingen zijn uiteenlopend en zijn samengevat in tabel 15. Daaruit blijkt over de PGA-elementen heen dat de feedback opgeteld voor 48% bestaat uit een algemene toelichtingen van de gemaakte keuze, 38% gaat over dat er een andere associatie is met de getoonde PGA-elementen Ook geeft 14% aan dat beide PGA-elementen niet intuïtief zijn (i.e. er is geen representatievoorkeur voor de getoonde PGA-elementen en de bijbehorende definitie).

Tabel 15: Samenvatting toelichting op de PGA-elementen

Categorie	Customer goal	Value proposition	Competence	Internal goal	Value statement	% totaal opgeteld
De toelichting geeft een algemene toelichting van de keuze	13	19	9	15	17	48%
De toelichting geeft aan dat er een andere associatie is met 3de getoonde PGA-elementen	16	8	11	10	12	38%
De toelichting geeft aan dat beide PGA-element niet intuïtief zijn	3	3	8	5	2	14%

In tabel 16 worden de resultaten weergegeven voor de persoonlijke resultaten en in tabel 17 worden de groepsresultaten weergegeven. De gevolgde testprocedure staat in bijlage O beschreven.

De bijbehorende hypothesen zijn als volgt:

- H₀: de representatievoorkeur van de nieuwe PGA-notatie voor element x tijdens taak y is gelijk aan die van de originele PGA-notatie voor element x tijdens taak y.
- H_{representatievoorkeur}: de representatievoorkeur van de nieuwe PGA-notatie voor element x is hoger dan die van de originele PGA-notatie voor element x tijdens taak y.

Als de toets aangeeft dat er een significant verschil is (e.g. er is meer aan de hand dan toeval) dan wordt de alternatieve hypothese (H_{representatievoorkeur}) aanvaard en de nulhypothese (H₀) wordt verworpen. Een significant verschil is als de p-value <= 0.05 is. Wanneer de toets geen significant verschil aangeeft dan wordt de alternatieve hypothese (H_{representatievoorkeur}) niet aanvaard en blijft de nulhypothese (H₀) niet verworpen.

De uitkomsten van de persoonlijke resultaten in tabel 16 voor de taak representatievoorkeursvragen laat zien dat er een significant verschil is voor de afhankelijke variabelen *customer goal*, *value proposition* en *internal goal*. Voor de afhankelijke variabele *competence* en *value stream* is het resultaat niet significant.

Tabel 16: Persoonlijke resultaten: uitkomsten representatievoorkeursvragen

Taak (y)	Afhankelijke variabele (x)	Label survey	Toets	Mean Rank (0/1)	Asymp. Sig. (2-tailed)	Sig. (1-tailed)	H ₀	H _{representatie-voorkeur}	Bron
Representatie-voorkeursvragen	<i>Customer goal</i>	CustomerGoalCode	One-Sample Wilcoxon Signed Rank Test	78.00	.004	.002	Verwerpen	Aanvaarden	Bijlage K
Representatie-voorkeursvragen	<i>Value proposition</i>	ValuePropositionCode	One-Sample Wilcoxon Signed Rank Test	73.00	.005	.003	Verwerpen	Aanvaarden	Bijlage K
Representatie-voorkeursvragen	<i>Competence</i>	CompetenceCode	One-Sample Wilcoxon Signed Rank Test	50.00	.260	0.87	Niet verwerpen	Niet aanvaarden	ijlage K
Representatie-voorkeursvragen	<i>Internal goal</i>	InternalGoalCode	One-Sample Test	53.21	.000	.000	Verwerpen	Aanvaarden	Bijlage K
Representatie-voorkeursvragen	<i>Value stream</i>	ValueStreamCode	One-Sample Test	41.43	.001	.9995	Niet verwerpen	Niet aanvaarden	Bijlage K

De uitkomsten van de groepsresultaten in tabel 17 voor de taak representatievoorkeursvragen laat zien dat er een significant verschil is voor de afhankelijke variabelen *customer goal* en *value proposition*. Voor de afhankelijke variabele *competence*, *internal goal* en *value stream* is het resultaat niet significant.

Tabel 17: Groepsresultaten: uitkomsten representatievoorkeursvragen

Taak (y)	Afhankelijke variabele (x)	Label survey	Toets	Mean Rank (0/1)	Asymp. Sig. (2-tailed)	Sig. (1-tailed)	H ₀	H _{representatie-voorkeur}	Bron
Representatie-voorkeursvragen	<i>Customer goal</i>	CustomerGoalCode	One-Sample Wilcoxon Signed Rank Test	63	.042	.021	Verwerpen	Aanvaarden	Bijlage L
Representatie-voorkeursvragen	<i>Value proposition</i>	ValuePropositionCode	One-Sample Wilcoxon Signed Rank Test	70	.001	.000	Verwerpen	Aanvaarden	Bijlage L
Representatie-voorkeursvragen	<i>Competence</i>	CompetenceCode	One-Sample Wilcoxon Signed Rank Test	50	.269	.866	Niet verwerpen	Niet aanvaarden	Bijlage L
Representatie-voorkeursvragen	<i>Internal goal</i>	InternalGoalCode	One-Sample Test	53.52	.406	.203	Niet verwerpen	Niet aanvaarden	Bijlage L
Representatie-voorkeursvragen	<i>Value stream</i>	ValueStreamCode	One-Sample Wilcoxon Signed Rank Test	50	.605	.698	Niet verwerpen	Niet aanvaarden	Bijlage L

4.5. Uitkomsten: geaggregeerd per PGA-element

De uitkomsten van de metingen over interpretatieve effectiviteit (zie §4.2) en representatievoorkeur (zie §4.4) zijn geaggregeerd (i.e. resultaten samenvoegen) per PGA-Element om de intuïtieve voorkeur per PGA-element te kunnen meten. De scores voor begripsvragen, notatie-associatievragen en representatievoorkeursvragen zijn herschaald om de analyse in SPSS te vereenvoudigen. De geaggregeerde scores zijn in percentages uitgedrukt per taak en opgeteld per PGA-element, zo ontstaat de geaggregeerde score per PGA-element. In bijlage O is de procedure uitgewerkt om tot een geaggregeerde score te komen.

In tabel 18 worden de resultaten weergegeven voor de persoonlijke resultaten en in tabel 19 zijn de groepsresultaten weergegeven.

De bijbehorende hypothesen zijn als volgt:

- H_0 : de geaggregeerde score van de nieuwe PGA-notatie voor element x is gelijk aan die van de originele PGA-notatie voor element x.
- $H_{\text{geaggregeerd}}$: de geaggregeerde score van de nieuwe PGA-notatie voor element x is hoger dan die van de originele PGA-notatie voor element x.

Als de toets aangeeft dat er een significant verschil is (e.g. er is meer aan de hand dan toeval) dan wordt de alternatieve hypothese ($H_{\text{geaggregeerd}}$) aanvaard en de nulhypothese (H_0) wordt verworpen. Een significant verschil is als de p-waarde ≤ 0.05 is. Wanneer de toets geen significant verschil aangeeft dan wordt de alternatieve hypothese ($H_{\text{geaggregeerd}}$) niet aanvaard en blijft de nulhypothese (H_0) niet verworpen.

De uitkomsten van de persoonlijke resultaten in tabel 18 voor de geaggregeerde score laat zien dat er geen significant verschil is voor de afhankelijke variabelen *customer goal*, *value proposition*, *competence*, *internal goal* en *value stream*.

Tabel 18: Persoonlijke resultaten: geaggregeerde elementen

Afhankelijke variabele (x)	Label survey	Toets	Mean Rank (0/1)	Asymp. Sig. (2-tailed)	Sig. (1-tailed)	H_0	$H_{\text{geaggregeerd}}$	Bron
<i>Customer goal</i>	GeaggregeerdCustomerGoal	Independent Samples Test	259.44/248.07	.737	.632	Niet verwerpen	Niet aanvaarden	Bijlage M
<i>Value proposition</i>	GeaggregeerdValueProposition	independent Samples Test	210.95/199.36	.793	.604	Niet verwerpen	Niet aanvaarden	Bijlage M
<i>Competence</i>	GeaggregeerdCompetence	Independent Samples Test	239.68/191.59	.139	.070	Niet verwerpen	Niet aanvaarden	Bijlage M
<i>Internal goal</i>	GeaggregeerdInternalGoal	Independent Samples Test	261.43/277.22	.526	.737	Niet verwerpen	Niet aanvaarden	Bijlage M
<i>Value stream</i>	GeaggregeerdValueStream	Mann-Whitney U test	8.79/6.21	.250	.875	Niet verwerpen	Niet aanvaarden	Bijlage M

De uitkomsten van de groepsresultaten in tabel 19 voor de geaggregeerde score laat zien dat er geen significant verschil is voor de afhankelijke variabelen *customer goal*, *value proposition*, *competence*, *internal goal* en *value stream*.

Tabel 19: Groepsresultaten: geaggregeerde elementen

Afhankelijke variabele (x)	Label survey	Toets	Mean Rank (0/1)	Asymp. Sig. (2-tailed)	Sig. (1-tailed)	H ₀	H _{geaggregeerd}	Bron
Customer goal	GeaggregeerdCustomerGoal	Mann-Whitney U test	23.08/21.80	.741	.630	Niet verwerpen	Niet aanvaarden	Bijlage N
Value proposition	GeaggregeerdValueProposition	Independent Samples Test	201.74/180.17	.367	.817	Niet verwerpen	Niet aanvaarden	Bijlage N
Competence	GeaggregeerdCompetence	Independent Samples Test	214.56/195.89	.394	.197	Niet verwerpen	Niet aanvaarden	Bijlage N
Internal goal	GeaggregeerdInternalGoal	Mann-Whitney U test	21.79/23.35	.689	.656	Niet verwerpen	Niet aanvaarden	Bijlage N
Value stream	GeaggregeerdValueStream	Mann-Whitney U test	24.73/19.83	.207	.897	Niet verwerpen	Niet aanvaarden	Bijlage N

4.6. Conclusie

In deze paragraaf zijn de uitkomsten van de metingen over interpretatieve effectiviteit (zie §4.2), interpretatieve efficiëntie (zie §4.3), representatievoorkeur (zie §4.4) en geaggregeerd per element (zie §4.5) samengevoegd in tabel 20 en tabel 21.

De persoonlijke resultaten in tabel 20 voor de hypothesen H_{effectiviteit}, H_{efficiëntie} en H_{geaggregeerd} zijn niet aanvaard voor de PGA-elementen *customer goal*, *value proposition*, *competence*, *internal goal* en *value stream*. Voor H_{representatievoorkeur} is de hypothese aanvaard voor de PGA-elementen *customer goal*, *value proposition* en *internal goal*. Voor de PGA-elementen *competence*, en *value steam* is de hypothese H_{representatievoorkeur} niet aanvaard.

Tabel 20: Persoonlijke resultaten: uitkomsten onderzoek

PGA-element / Taak	Hypothese							
	H _{effectiviteit}			H _{efficiëntie}			H _{representatievoorkeur}	H _{geaggregeerd}
	Notatie-associatievragen	Begripsvragen	Herinnerings-Vragen	Notatie-associatievragen	Begripsvragen	Herinnerings-Vragen	Representatie-voorkeursvragen	Niet aanvaarden
Customer goal	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Aanvaarden	Niet aanvaarden
Value proposition	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Aanvaarden	Niet aanvaarden
Competence	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden
Internal goal	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Aanvaarden	Niet aanvaarden
Value stream	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden

De groepsresultaten in tabel 21 laten zien dat voor het PGA-element *customer goal* de hypothese H_{representatievoorkeur} bij representatievoorkeursvragen is aanvaard. Voor het PGA-element *value proposition* bij notatie-associatievragen is de hypothese H_{effectiviteit} aanvaard en bij de representatievoorkeursvragen is de hypothese H_{representatievoorkeur} aanvaard. Voor het PGA-element *competence* is voor de herinneringsvragen de hypothese H_{efficiëntie} aanvaard en voor het PGA-element *internal goal* is voor de herinneringsvragen de hypothese H_{efficiëntie} aanvaard. Bij de overige vragen zijn de bijbehorende hypothesen niet aanvaard.

Tabel 21: Groepsresultaten: uitkomsten onderzoek

PGA-element	Hypothese							
	H _{effectiviteit}			H _{efficiëntie}			H _{representatievoorkeur}	H _{geaggregeerd}
	Notatie-associatievragen	Begripsvragen	Herinnerings-Vragen	Notatie-associatievragen	Begripsvragen	Herinnerings-Vragen	Representatie-voorkeursvragen	Niet aanvaarden
<i>Customer goal</i>	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Aanvaarden	Niet aanvaarden
<i>Value proposition</i>	Aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Aanvaarden	Niet aanvaarden
<i>Competence</i>	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Aanvaarden	Niet aanvaarden	Niet aanvaarden
<i>Internal goal</i>	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Aanvaarden	Niet aanvaarden	Niet aanvaarden
<i>Value stream</i>	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden*	Niet aanvaarden	Niet aanvaarden	Niet aanvaarden

*Bij het beantwoorden van begripsvragen zijn de PGA-elementen twee keer uitgevraagd. De groepsresultaten voor begripsvragen en het element value stream hebben geen eenduidig resultaat opgeleverd (zie §4.3).

De uitkomsten van de metingen interpretatieve effectiviteit (i.e. H_{effectiviteit}) en representatievoorkeur (i.e. H_{representatievoorkeur}) zijn samengevoegd (i.e. H_{geaggregeerd}). De geaggregeerde resultaten laten zien dat voor de elementen *customer goal*, *value proposition*, *competence*, *internal goal* en *value stream* er geen intuïtieve voorkeur is voor de nieuwe PGA-elementen voor zowel de persoonlijke resultaten als de groepsresultaten. De conclusie is dat de uitkomsten niet aantonen dat de nieuwe versie van de PGA-notatie intuïtiever is dan de oorspronkelijke versie. Ook is niet aangetoond dat, wanneer de nieuwe PGA-elementen gebruikt worden bij het beantwoorden van notatie-associatievragen, begripsvragen en herinneringsvragen deze efficiënter worden beantwoord.

Het antwoord op deelvraag drie is dat de persoonlijke resultaten en groepsresultaten gezamenlijk niet aantonen dat de nieuwe PGA-notatie intuïtiever is dan de oorspronkelijke PGA-notatie.

5. Discussie, conclusies en aanbevelingen

Dit hoofdstuk beantwoordt de onderzoeksvraag, de verklaring van het onderzoek en geeft aanbevelingen voor verder onderzoek.

5.1. Discussie – reflectie

De voorgestelde aanpassingen aan de PGA-notatie uit het artikel van Roelens en Bork (2020) zijn in dit onderzoek geëvalueerd. Er had immers nog geen evaluatie plaatsgevonden om de intuïtiviteit van de nieuwe PGA-notatie ook effectief aan te tonen. De resultaten in dit onderzoek laten zien dat de voorgestelde PGA-verbeteringen aan de intuïtiviteit niet significant zijn voor potentiële eindgebruikers.

Vooraf aan het onderzoek waren de verwachtingen dat enkele nieuwe PGA-elementen intuïtiever zouden zijn dan de oorspronkelijke versie. De uitkomsten van de metingen over interpretatieve effectiviteit en representatievoorkeur zijn geaggregeerd (i.e. resultaten samenvoegen) per PGA-element om de intuïtieve voorkeur per PGA-element te kunnen meten. Voor de nieuwe PGA-elementen is het niet aangetoond dat deze intuïtiever zijn dan de oorspronkelijke versie (zie §4.6). De uitkomsten komen daarmee niet overeen met de verwachting van de resultaten op basis van het literatuuronderzoek in hoofdstuk 2 en de uitwerking van de aanpak in hoofdstuk 3.

De uitkomsten over de hypothese representatievoorkeur (zie §4.6) komen meer overeen met de verwachting dat deze PGA-elementen intuïtiever zijn dan de huidige. Voor de elementen *customer goal* en *value proposition* is een duidelijke voorkeur aangegeven voor zowel de persoonlijke resultaten als de groepsresultaten voor de nieuwe PGA-notatie. Er is een verschil te zien in de resultaten tussen de taken (e.g. notatie-associatievragen, begripsvragen, herinneringsvragen) die horen bij effectiviteit ten opzichte van de representatievoorkeursvragen. Het verschil zit in het feit dat in de taken die horen bij effectiviteit de participanten hebben gewerkt met één notatie per PGA-element en bij representatievoorkeursvragen is nadrukkelijk gevraagd om een voorkeur aan te geven tussen twee PGA-elementen, inclusief de toelichting.

De toelichtingen bij de representatievoorkeursvragen zijn samengevat in tabel 15. Daaruit blijkt dat 48% van de toelichtingen gaan over algemene opmerkingen c.q. bevestigingen van gemaakte keuzes, 38% geeft in de toelichting aan dat er een andere associatie is met één van de getoonde PGA-elementen en 14% geeft aan dat beide PGA-elementen niet intuïtief zijn. De toelichtingen zijn een belangrijke bron die gebruikt kunnen worden in het vervolgonderzoek om de intuïtiviteit van de PGA-elementen verder te optimaliseren.

De data van de persoonlijke resultaten is geverifieerd. In totaal hebben 18 participanten een survey ingevuld en zijn vier resultaten ongeldig verklaard, omdat deze niet volledig ingevuld waren. De persoonlijke resultaten bestaan uit de dataset van 14 participanten en voor de groepsresultaten bestaat de dataset uit 44 participanten. De meetmethodes, zoals beschreven in §3.2, zijn toegepast om de afhankelijke variabelen te meten. Door gebruik te maken van een dubbele analyse van de persoonlijke resultaten en groepsresultaten zijn de hypothesen (e.g. $H_{\text{effectiviteit}}$, $H_{\text{efficiëntie}}$, $H_{\text{representatievoorkeur}}$, en $H_{\text{geaggregeerd}}$) statistisch getoetst. In het onderzoek zijn de persoonlijke resultaten vergeleken met de groepsresultaten (zie §4.6). Op basis van de uitkomsten van de geaggregeerde resultaten kunnen voor, zowel de persoonlijke resultaten als groepsresultaten dezelfde conclusie worden getrokken. Om verstoringen uit te sluiten zijn dezelfde taken uitgevoerd voor het evalueren van de originele PGA-notatie en de nieuwe PGA-notatie. Om onder dezelfde omstandigheden de taken uit te voeren is gebruik gemaakt van de tool LimeSurvey, zodat het gemeten effect niet verstoord wordt. Deze tool heeft geholpen om de tijd op een betrouwbare manier te meten.

Als casus is gekozen voor een bedrijf dat een toonaangevende marktplaats biedt voor online voedsellevering. Daarbij is gesteld dat de participanten achtergrondkennis nodig hebben om in staat te zijn om strategische bedrijfsprocessen in te kunnen schatten, zonder dat daarvoor inhoudelijke kennis aanwezig is. De respondenten zijn niet bekend met de PGA-methode. In de persoonlijke resultaten heeft 64% enige kennis van modelleringstechnieken en bij de groepsresultaten is dat 59%.

De resultaten uit dit onderzoek kunnen worden gegeneraliseerd op het gebied van de intuïtiviteit van de PGA-notatie door het in kaart brengen van strategische bedrijfsprocessen in een logistieke keten. Triangulatie is toegepast, doordat er gebruikt gemaakt is van twee databronnen (e.g. persoonlijke resultaten en groepsresultaten) en bij de representatievoorkeurstaak is er ook nadrukkelijk nog om aanvullende informatie gevraagd. Er is dus vanuit verschillende invalshoeken onderzoek gedaan om de onderzoeksvraag te kunnen beantwoorden. Het resultaat heeft een onderzoeksmethode opgeleverd (zie §3.2) die toegepast is in dit onderzoek. In bijlage D zijn de links naar de origineel bronbestanden terug te vinden. Voordat de participanten konden beginnen met het invullen van de survey hebben de participanten eerst een verklaring ingevuld voor deelname aan het wetenschappelijk onderzoek. Daarin is o.a. beschreven dat het onderzoek op elk moment beëindigd kan worden, zonder opgave van reden. Ook zijn de participanten vooraf geïnformeerd over het doel van het onderzoek en is nadrukkelijk aangegeven dat het onderzoek anoniem wordt uitgevoerd.

5.2. Conclusies

Dit onderzoek geeft antwoord op de overkoepelende onderzoeksvraag: “Hoe kan de intuïtiviteit van twee verschillende versies van de PGA-notatie geëvalueerd worden door het uitvoeren van een experiment binnen een praktische bedrijfscontext?”. Hiervoor is kwantitatief experimenteel onderzoek uitgevoerd naar het effect op de intuïtiviteit van de twee verschillende PGA-notatie versies.

De resultaten van het literatuuronderzoek (i.e. deelonderzoeksvraag 1) laten zien dat er verschillende methodes zijn toegepast om conceptuele modelleertalen empirisch te evalueren. Op basis van de afhankelijke, onafhankelijke variabelen en de metingen, blijkt dat er geen eenduidige methode bestaat voor het evalueren van conceptuele modellen. Er zitten immers verschillen in de gebruikte variabelen en de metingen. Op basis van de resultaten van de doeltreffendheid (zie tabel 6) blijkt dat begripsvragen het meeste voorkomen en herinneringsvragen, probleemoplossende vragen, en invulvragen minder zijn toegepast.

Een deductieve aanpak is gekozen voor het beschrijven van de experimentele methode die bijdraagt bij het evalueren van de intuïtiviteit van de PGA-modelleertaal (i.e. deelonderzoeksvraag 2). Het experiment onderzoekt of de onafhankelijke variabelen (i.e. oorspronkelijke PGA-notatie, nieuwe PGA-notatie) van invloed kunnen zijn op de afhankelijke (i.e. interpretatieve effectiviteit, interpretatieve efficiëntie en intuïtieve begrijpelijkheid). De conclusie is dat notatie-associatievragen, begripsvragen, herinneringsvragen en representatievoorkeur onderdeel zijn van het experiment om de onafhankelijke variabelen te meten. Hierdoor kan de intuïtiviteit van de PGA-modelleertaal worden geëvalueerd.

De resultaten van het experiment (i.e. deelonderzoeksvraag 3) over de meting interpretatieve effectiviteit en representatievoorkeur zijn geaggregeerd (i.e. resultaten samenvoegen) per PGA-element om de intuïtieve voorkeur per PGA-element te kunnen meten. Voor de nieuwe PGA-elementen is het niet aangetoond dat deze intuïtiever zijn dan de oorspronkelijke versie (zie §4.6). De conclusie is dat voor de elementen *customer goal*, *value proposition*, *competence*, *internal goal* en *valuestream* er geen intuïtieve voorkeur is voor de nieuwe PGA-elementen.

In relatie tot de overkoepelende onderzoeksvraag is gekozen voor het uitvoeren van een experiment binnen een praktische bedrijfscontext door op basis van de literatuur een experimentele methode op te stellen om de intuïtiviteit tussen twee verschillende versies PGA-modelleertaal te evalueren. Daaruit blijkt dat de oorspronkelijke methodes niet voorzien in het meten van de intuïtiviteit tussen twee verschillende versies van conceptuele modelleertalen en deze te evalueren. Om de intuïtieve begrijpelijkheid te meten is het experiment uitgebreid met het meten van de representatievoorkeur. De conclusie is dat de uitkomsten van het experiment niet aantonen dat de nieuwe versie van de PGA-notatie intuïtiever is dan de oorspronkelijke versie.

5.3. Aanbevelingen voor de praktijk

Uit het onderzoek blijkt dat de visuele wijzigingen aan de nieuwe PGA-elementen onvoldoende zijn om aan te tonen dat de nieuwe PGA-elementen intuïtiever zijn dan de oorspronkelijke PGA-elementen. De resultaten van de representatievoorkeursvragen hebben aangetoond dat voor de PGA-elementen *customer goal* en *value proposition* de voorkeur is voor de nieuwe PGA-notatie. In combinatie met de notatie-associatievragen, begripsvragen en herinneringsvragen is dit niet significant gebleken. Deze resultaten van de representatievoorkeursvragen en de bijbehorende toelichting geven aanleiding om de vijf PGA-elementen (e.g. *customer goal*, *value proposition*, *competence*, *internal goal* en *value stream*) te herontwerpen door een expert op het gebied van grafische vormgeving. Vervolgens kan er op basis van het onderzoek van Roelens en Bork (2020) een experimentele evaluatie worden uitgevoerd over de nieuwe vormgegeven PGA-elementen. Na het verwerken van de uitkomsten kunnen de nieuwe vormgegeven PGA-elementen op basis van deze onderzoeks aanpak worden geëvalueerd met de oorspronkelijke versie.

5.4. Aanbevelingen voor verder onderzoek

Om de intuïtieve begrijpelijkheid te meten is gebruik gemaakt van notatie-associatievragen, begripsvragen, herinneringsvragen en representatievoorkeursvragen. Uit het onderzoek blijkt dat de uitkomsten van representatievoorkeursvragen een belangrijke rol spelen om de intuïtieve voorkeur te bepalen tussen de twee versies. Door het uitvoeren van eerst de notatie-associatievragen, begripsvragen, herinneringsvragen en vervolgens de representatievoorkeursvragen is er een zekere mate van voorkennis ontstaan, wanneer naar de volgende taak wordt gegaan. Dit heeft mogelijk invloed op de intuïtieve begrijpelijkheid (i.e. het gemak waarbij het PGA-element zonder enige voorkennis of opleiding onmiddellijk door de participanten kan worden begrepen). Een aanbeveling voor verder onderzoek zou kunnen zijn om te onderzoeken wat het effect is op de intuïtieve begrijpelijkheid om representatievoorkeursvragen te combineren met notatie-associatievragen, begripsvragen en herinneringsvragen. Hiervoor kan een experiment worden opgezet om de uitkomsten van alleen representatievoorkeursvragen te vergelijken met een gecombineerde aanpak met notatie-associatievragen, begripsvragen, herinneringsvragen en representatievoorkeursvragen.

Referenties

- Bork, D., Schrüffer, C., & Karagiannis, D. (2019). Intuitive Understanding of Domain-Specific Modeling Languages: Proposition and Application of an Evaluation Technique. In (pp. 311-319).
- Burton-Jones, A., & Meso, P. N. (2006). Conceptualizing Systems for Understanding: An Empirical Test of Decomposition Principles in Object-Oriented Analysis. *Information Systems Research*, 17(1), 38-60.
- Burton-Jones, A., Wand, Y., & Weber, R. (2009). Guidelines for Empirical Evaluations of Conceptual Modeling Grammars. *JOURNAL OF THE ASSOCIATION FOR INFORMATION SYSTEMS*, 10(6), 495-532.
- Cruzes, D. S., Vennesland, A., & Natvig, M. K. (2013). *Empirical Evaluation of the Quality of Conceptual Models Based on User Perceptions: A Case Study in the Transport Domain*, Berlin, Heidelberg.
- Gemino, A., & Wand, Y. (2004). A framework for empirical evaluation of conceptual modeling techniques. *Requirements Engineering*, 9(4), 248-260.
- Houy, C., Fettke, P., & Loos, P. (2012). *Understanding Understandability of Conceptual Models – What Are We Actually Talking about?*, Berlin, Heidelberg.
- Jošt, G., Huber, J., Heričko, M., & Polančič, G. (2016). An empirical investigation of intuitive understandability of process diagrams. *Computer Standards & Interfaces*, 48, 90-111.
- Lankhorst, M. (2017). *Enterprise Architecture at Work: Modelling, Communication and Analysis*: Springer Publishing Company, Incorporated.
- Luoma, J., Kelly, S., & Tolvanen, J. (2004). Defining Domain-Specific Modeling Languages: Collected Experiences.
- Maes, A., & Poels, G. (2007). Evaluating quality of conceptual modelling scripts based on user perceptions. *Data & Knowledge Engineering*, 63(3), 701-724.
- Mending, J., Recker, J., Reijers, H. A., & Leopold, H. (2018). An Empirical Review of the Connection Between Model Viewer Characteristics and the Comprehension of Conceptual Process Models. *Information systems frontiers*, 21(5), 1111-1135.
- Moody, D. (2010). The “Physics” of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering. *Software Engineering, IEEE Transactions on*, 35, 756-779. doi:10.1109/TSE.2009.67
- Oberconsulting. (2020). Essenties van Bedrijfsarchitectuur. Retrieved from <https://www.oberconsulting.nl/bedrijfsarchitectuur/9-essenties-van-bedrijfsarchitectuur>
- OMG. (2014). Business Process model and Notation. Retrieved from <https://www.omg.org/spec/BPMN/>
- OpenGroup. (2018). *The TOGAF standard, version 9.2*: Van Haren Publishing.
- OpenGroup. (2019). *ArchiMate® 3.1 Specification*: Van Haren Publishing.
- Realdolmen. (2016). Retrieved from <https://www.realdolmen.com/sites/default/files/2016-09/WhitePaper-Bedrijfsstrategie%20als%20basis%20voor%20strategische%20ICT-keuzes-nl.pdf>
- Recker, J., Rosemann, M., Green, P., & Indulska, M. (2011). Do Ontological Deficiencies in Modeling Grammars Matter? *MIS Quarterly*, 35(1), 57-79.
- Roelens, B., & Bork, D. (2020). Performing an experimental evaluation to improve the initial PGA notation.
- Roelens, B., Steenacker, W., & Poels, G. (2019). Realizing strategic fit within the business architecture: the design of a Process-Goal Alignment modeling and analysis technique. *Software & Systems Modeling*, 18(1), 631-662.
- Saunders, M. N. K., Lewis, P., & Thornhill, A. (2019). *Research Methods for Business Students*.
- Schraivesande, M. (2010). Bedrijfsarchitectuur draagt bij aan duurzaam ondernemen. Retrieved from <http://vianovaarchitectura.nl/page/bedrijfsarchitectuur-draagt-bij-aan-duurzaam-ondernemen>

Shanks, G., Tansley, E., Nuredini, J., Tobin, D., & Weber, R. (2008). Representing Part-Whole Relations in Conceptual Modeling: An Empirical Evaluation. *MIS Quarterly*, 32(3), 553-573.

Versnellingsplan. (2020). Toetsmatrix. Retrieved from <https://sh-studiedata.nl/R/Toetsmatrix.html>

Zachman, J. A. (1982). Business Systems Planning and Business Information Control Study: A comparison. *IBM Systems Journal*, 21(1), 31-53.

Bijlage A: Resultaten building blocks methode

In tabel 22 zijn de referenties opgenomen die tijdens de *building blocks* methode voor de zoektermen “*conceptual modeling*” en “*empirical evaluation*” zijn geraadpleegd.

Tabel 22: Geselecteerde artikelen *building blocks* methode

Nr.	Referentie	Titel	Jaar	Context	Respondent	Type vraag
1	Shanks et al. (2008)	<i>Representing part-whole relations in conceptual modeling: An empirical evaluation.</i>	2008	Ja	Business expert	Problem-solving
2	Gemino en Wand (2004)	<i>A framework for empirical evaluation of conceptual modeling techniques.</i>	2004	Ja	Business expert	Recal, Comprehension, Problem-solving, Cloze
3	Recker et al. (2011)	<i>Do ontological deficiencies in modeling grammars matter?</i>	2011	Ja	Business expert	Comprehension
4	Bera et al. (2014)	<i>How semantics and pragmatics interact in understanding conceptual models.</i>	2014	Ja	Studenten	Comprehension, Problem-solving
5	Bera et al. (2014)	<i>Guidelines for using UML association classes and their effect on domain understanding in requirements engineering.</i>	2014	Ja	Studenten	Problem-solving
6	Allen et al. (2012)	<i>A research note on representing part-whole relations in conceptual modeling.</i>	2012	Ja	Studenten	Comprehension
7	Burton-Jones et al. (2009)	<i>Guidelines for Empirical Evaluations of Conceptual Modeling Grammars.</i>	2009	Ja	Business expert, studenten	Comprehension, Problem-solving
8	Hutchison et al. (2008)	<i>Conceptual modeling: ER 2008: 27th international conference on conceptual modeling.</i>	2008	Nee	-	-
9	Gailly et al. (2017)	<i>Recommendation-Based Conceptual Modeling and Ontology Evolution Framework (CMOE+).</i>	2017	Ja	Studenten	Comprehension
10	N.A (2013)	<i>Conceptual Modeling: 32th International Conference, ER 2013, Hong-Kong, China, November 11-13, 2013. Proceedings.*</i> <i>Publicatie van Cruzes, D. S., Vennesland, A., & Natvig, M. K. (2013). Empirical Evaluation of the Quality of Conceptual Models Based on User Perceptions: A Case Study in the Transport Domain, Berlin, Heidelberg.</i>	2013	Ja	Business expert	Comprehension
11	Hutchison et al. (2011)	<i>Conceptual Modeling - ER 2011: 30th International Conference, ER 2011, Brussels, Belgium, October 31 - November 3, 2011. Proceedings.*</i>	2011	Nee	-	-
12	Ng et al. (2013)	<i>Conceptual Modeling - ER 2013: 32th International Conference, ER 2013 Hong-Kong, China, November 11-13, 2013, Proceedings.*</i> <i>Met uitzondering van de publicatie van Cruzes et al. (2013) die is beoordeeld op nummer 10.</i>	2013	Nee	-	-

*Deze publicaties zijn beoordeeld door te zoeken op “*empirical evaluation*” om zo kandidaatartikelen te vinden en te beoordelen op basis van het selectiecriterium ‘context’.

Bijlage B: Resultaten forward snowballing methode

In tabel 23 zijn de referenties opgenomen die tijdens *de forward snowballing* methode zijn geraadpleegd op basis van het artikel van Gemino en Wand (2004).

Tabel 23: Geselecteerde artikelen snowballing methode

Nr.	Referentie	Titel	Jaar	Context	Respondent	Type vraag
1	Figl et al. (2017)	Comprehension of procedural visual business process models: A literature review.	2017	Nee	Business expert, studenten	Comprehension
2	Figl et al. (2017)	Exploring cognitive style and task-specific preferences for process representations.	2016	Ja	Studenten	Comprehension
3	Mending et al. (2018)	An Empirical Review of the Connection Between Model Viewer Characteristics and the Comprehension of Conceptual Process Models.	2018	Ja	B	Comprehension, Knowledge
4	Burton-Jones et al. (2006)	Conceptualizing Systems for Understanding: An Empirical Test of Decomposition Principles in Object-Oriented Analysis.	2006	Ja	Studenten	Comprehension, Problem-solving
5	Recker (2013)	Empirical investigation of the usefulness of Gateway constructs in process models.	2013	Geen toegang tot het artikel via de OU bibliotheek.	-	-
6	Recker (2010)	Continued use of process modeling grammars: the impact of individual difference factors.	2010	Geen toegang tot het artikel via de OU bibliotheek.	-	-
7	Khatri et al. (2014)	Capturing Telic/Atelic Temporal Data Semantics: Generalizing Conventional Conceptual Models.	2014	Geen toegang tot het artikel via de OU bibliotheek.	-	-
8	Nassour et al. (2019)	Evaluating the comprehension of means-ends maps.	2019	Nee	Studenten	Comprehension
9	Oppl (2017)	Supporting the Collaborative Construction of a Shared Understanding About Work with a Guided Conceptual Modeling Technique.	2017	Nee	Studenten	Comprehension

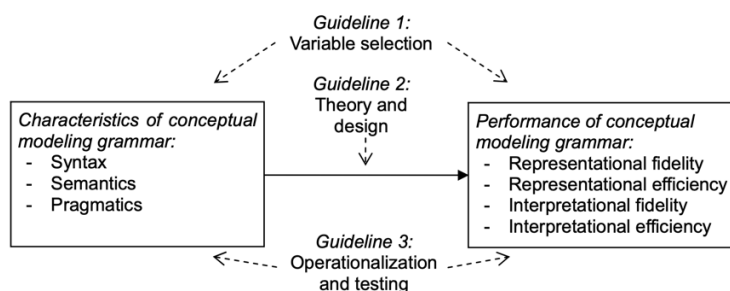
Bijlage C: Samenvatting literatuuronderzoek

In deze bijlage zijn de samenvattingen van de artikelen toegevoegd die onderdeel zijn van het theoretisch kader in relatie tot het empirisch evalueren van conceptuele modellen. Het beschrijft in het kort het doel van het artikel, de onderzoeksvraag, gebruikt materiaal, type onderzoek, aantal participanten en de gevolgde procedure. Hiervan kan per artikel afgeweken zijn, omdat een aantal artikelen gaan over het beschrijven van een *framework* of richtlijn rondom het empirisch evalueren van conceptuele modellen.

H1 Artikel: Burton-Jones et al. (2009)

Burton-Jones et al. (2009) beschrijven in het artikel dat er verschillende richtlijnen zijn ontwikkeld om onderzoekers te helpen met het evalueren van de grammatica via conceptuele modellen (Wand & Weber, 2002), (Gemino & Wand, 2004), (Siau, 2004), (Parsons & Cole, 2005), (Aguirre-Urreta & Marakas, 2008). Dit artikel geeft een verduidelijking van de richtlijnen met betrekking tot: (1) criteria voor de evaluatie van grammaticaprestaties, (2) kenmerken die de grammaticaprestaties kunnen beïnvloeden en (3) factoren waarmee rekening gehouden moet worden bij het testen van het effect van de kenmerken van de grammatica op de prestaties (Burton-Jones & Meso, 2006). Het artikel draagt bij aan het geven van richtlijnen om conceptuele modellen empirisch te evalueren, ongeacht de onderzoeksmethoden. Ook worden concrete richtlijnen verduidelijkt die gemakkelijk verkeerd kunnen worden geïnterpreteerd.

Burton-Jones et al. (2009) schrijven drie richtlijnen voor om conceptuele modelleringsgrammatica empirisch te evalueren op het gebied van effectiviteit en efficiëntie (figuur 3).



Figuur 3: Richtlijnen empirisch evalueren van conceptuele modelleringsgrammatica, overgenomen uit (Burton-Jones et al., 2009)

Richtlijn 1: Variabelen selecteren

Bij het evalueren van conceptuele modelleringsgrammatica wordt aangegeven dat onderzoekers baat hebben bij het begrip van het bereik van potentiële voorspellende variabelen en uitkomstvariabelen die gebruikt kunnen worden. De vier uitkomstvariabelen die gebruikt kunnen worden om de geproduceerde modellen te evalueren met behulp van conceptuele modelleringsgrammatica zijn:

1. *Representational fidelity*: hoe vertegenwoordigt iemand de perceptie van het model of de manier waarop een groep de betekenis van het domein heeft besproken?
2. *Representational efficiency*: welke middelen worden gebruikt om het model te maken?
3. *Interpretational fidelity*: hoe betrouwbaar is de betekenis van het model?
4. *Interpretational efficiency*: welke middelen worden gebruikt om het model te interpreteren?

Voorspellende variabelen kunnen van invloed zijn op de uitkomst variabelen. De focus van deze richtlijn zijn gericht op de grammatica die van invloed zijn op onderstaande variabelen:

1. Syntaxis: de concepten in de grammatica en de regels voor de indeling.
2. Semantics: de betekenis van de concepten in de grammatica.
3. Pragmatiek: de context waarin de grammatica wordt gebruikt.

Richtlijn 2: Theorie en ontwerp

Deze richtlijn beschrijft dat het een gebruikelijke manier is om op basis van een willekeurige steekproef van lezers, vragen te laten beantwoorden over de scrips. Een probleem zou kunnen zijn dat de respondenten niet antwoorden op basis van het model. Om dit te voorkomen is het belangrijk om achtergrondinformatie van het getoonde model en het domein te hebben. Onderzoekers kunnen dit op drie manieren aan de orde stellen: (1) vooraf aan het onderzoek wordt aan de participanten gevraagd of zij menen dat het model van belang is voor het uitvoeren van de taken, (2) na het uitvoeren van de taak wordt aan de participanten gevraagd om feedback om de herkenbaarheid van het model. Bij een lage herkenbaarheid van het model wordt deze uitgesloten van de analyse en (3) wanneer er geen verschillen te zien zijn tussen de verschillende behandelingsgroepen dan moet de beoordeling in twijfel worden getrokken of een andere verklaring gezocht worden. Dit kan bijvoorbeeld een slecht literatuuronderzoek en/of onderzoeksmethode zijn.

Richtlijn 3: Testen van de theorie

Elke test van de theorie vereist dat onderzoekers testen op de aanwezigheid van afwijkingen. Om de theorie toe te kunnen passen zullen onderzoekers een situatie moeten identificeren of creëren waarin meerdere grammatica's bestaan die variëren in het aantal defecten.

Bij het toetsen van de theorie wordt dan onderzocht of het aantal afwijkingen in de grammatica's (i.e. of in de modellen) samenhangt met de variantie in een van de uitkomstvariabelen.

H2 Artikel: Cruzes et al. (2013)

Cruzes et al. (2013) rapporteren in dit artikel over een empirische studie over het evalueren van de kwaliteit van een conceptueel model in het domein van de transportlogistiek, waarbij gebruik gemaakt wordt van bestaande kwaliteitsevaluatiekaders. Het doel is om de gebruikerstevredenheid van het framework te onderzoeken maar ook de uitdagingen, voordelen en verbeteringen in kaart te brengen. Het artikel bouwt met name voort op de vragen en metingen van Maes en Poels (Maes en Poels, 2007) met als doel om een kwalitatieve interviewhandleiding te creëren die de beoogde kwaliteit van een conceptueel *framework* verzamelt op basis van gebruikerservaring. Het gepresenteerde conceptuele *framework* (CF) heeft als doel om de verschillende stakeholdergroepen met verschillende kennisniveaus te laten samenwerken in het transport logistieke domein rond een gemeenschappelijke visie op de processen binnen het domein. Maes en Poels. (2007) evaluatie *framework* is gericht op gebruikersaspecten van de kwaliteit van het conceptuele model. De reden is dat het is afgestemd op de technologieën die worden gebruikt om deze technologie aan gebruikers uit te leggen en het evaluatie *framework* levert een gevalideerd meetinstrument dat uitsluitend gebaseerd is op conceptuele modevaluaties van eindgebruikers. Het meten van de gebruikersperceptie van de kwaliteit van het conceptueel model is een bepalende factor voor de tevredenheid van de gebruikers met het model.

Het kwaliteitsmodel bestaat uit vier componenten:

1. *The perceived semantic quality*: is een meting op hoe goed de beoogde kwaliteit wordt overgebracht en geïnterpreteerd door de ontvanger.
2. *The perceived ease of understanding*: is gebaseerd op hoe de gebruikers het model zullen beoordelen op hoe gemakkelijk het is om de informatie in het model te lezen en te kunnen begrijpen.
3. *The user satisfaction*: is de inhoudelijke evaluatie van de maten waarin gebruikers tevreden zijn over het model.
4. *The perceived usefulness*: is een concept voor het meten van de algemene kwaliteit van het model.

Het CF bestaat uit verschillende gezichtspunten die delen van de transport logistieke sector beschrijven, waarbij de volgende componenten een rol spelen en als materiaal gebruikt zijn:

- Een referentiemodel van de transport logistieke sector onderverdeeld in sub domeinen en betreft de volgende aspecten: vraag, aanbod, infrastructuur en regelgeving.
- Een beschrijving van rollen die relevant zijn voor elk sub-domein.
- Een set van activiteiten (i.e. *UML use case* modellen) en procesdiagrammen (e.g. BPNM) die relevant zijn voor de functies die door de rollen worden uitgevoerd.
- Een set aan informatiemodellen (i.e. *UML class* modellen) om de relatie te beschrijven tussen de sub-domeinen.

De data zijn verzameld door het houden van semigestructureerd interviews op basis van de Burton-Jones en Meso (2006) uitgangspunten. Het verschil in aanpak is dat Cruzes et al. (2013) een meer verkennende en kwalitatieve benadering hanteren. De aanpak is om de participanten te stimuleren om te praten over de problematiek rondom het CF. De zeven interviews zijn door twee participanten afgenomen en duurde ongeveer een uur per interview. De participanten werden geïnformeerd over het belang van het onderzoek en dat er geluidsopnames gemaakt werden tijdens het afnemen van het interview. De participanten waren verschillende soorten gebruikers van het framework en zijn op specifieke onderdelen van het *CF geïnterviewd*.

H3 Artikel: Gemino en Wand (2004)

Gemino en Wand (2004) presenteren een *framework* voor het empirisch evalueren van conceptuele modelleringstechnieken. Het *framework* is gebaseerd op het idee om modelleringstechnieken te vergelijken via de achterliggende grammatica. Het artikel bespreekt een theoretisch kader die als uitgangspunt dient voor het opstellen van een hypothese en het verkrijgen van de variabelen.

Een conceptueel model kent vier rollen in de ontwikkeling van het domeinbegrip: (1) ondersteuning van iemands eigen begripsvorming, (2) communiceren van details over het domein aan de *stakeholders*, (3) communiceren van domeingegevens aan systeemontwerpers en (4) het documenteren van het domein als referentie. Gemino en Wand (2004) beschrijven dat conceptuele modellering gezien kan worden als een proces waarbij individuen nadenken en communiceren over een domein met als doel om dit beter te begrijpen. Het artikel beschrijft een hoog foutpercentage in informatiesysteemprojecten en brengen dit in verband met gebrek aan input van de gebruiker, onvolledig of onduidelijke eisen en verandering van eisen. Conceptuele modelleringstechnieken dragen bij aan het formaliseren van het domein en om het proces van eisen beter te analyseren om zo een bij dragen te leveren aan het succes van het project.

Een conceptuele modelleertaal bestaat uit een set van concepten, vaak grafische symbolen en regels. De set van concepten en regels worden conceptuele modelleergrammatica genoemd. Het model is het resultaat van het conceptuele modelleerproces. Grammatica wordt gebruikt voor het maken van een beschrijving van een domein. Gemino en Wand (2004) beschrijven een gebrek aan empirische vergelijkingen van conceptuele modelleringstechnieken en introduceren een generiek *framework*. Het *framework* is bedoeld op empirisch werk te classificeren met betrekking tot de aspecten die onderzoekers kunnen kiezen bij het ontwerpen van experimenten en het meten van de resultaten.

Uit eerder studies is gebleken dat er verschillen zitten in de concepten die worden gebruikt bij het meten van de resultaten van de vergelijking van conceptuele modelleringstechnieken. Gemino en Wand (2004) benadrukken het onderscheid tussen het product van het proces en het proces zelf en beschrijven in het framework twee criteria voor de vergelijking: effectiviteit en efficiëntie.

Voor het meten van het model kunnen interpretatietaken worden toegepast. Dit is van toepassing op het interpreteren van een bestaand model en ligt minder voor de hand bij het creëren van het model. Een resultaat van interpreteren van het model is het cognitieve model die door de gebruiker is gemaakt. Dit model kan niet direct worden beoordeeld. Wat wel beoordeeld kan worden is het uitvoeren van taken op basis van kennis, zo zijn er verschillende metingen voor het vastleggen van het begrip dat een participant ontwikkelt bij het interpreteren van het model. Het interpreteren kan door het uitvoeren van een begripstest en bestaat meestal uit een reeks van begripvragen en een set probleemoplossende vragen over elementen die in het model in overweging worden genomen. Een andere test kan worden uitgevoerd wanneer het model beschikbaar is en aangeeft hoe het model functioneert of nadat het is verwijderd en hoe het wordt herinnerd. Een ander mogelijkheid om het resultaat te meten is een semantische *recall* op basis van een invulproef (i.e. *Close*). Deze test gaat over het invullen van lege plekken in stukken tekst nadat het model is verwijderd. Hoe meer er herinnerd wordt hoe groter het begrip van het domein.

H4 Artikel: Mendling et al. (2018)

Mendling et al. (2018) hebben onderzocht wat de factoren waren die van invloed zijn op de manier waarom personen de betekenis van de conceptuele modellen beïnvloeden, wat belangrijk is voor het verbeteren van het ontwerp van informatiesystemen in het kader van effectiviteit en efficiëntie. Het artikel bestudeert eerst hoe verschillende *model viewer characteristics* (i.e. MVC'S) zijn gebruikt in eerdere experimentele onderzoeken en hoe deze zich tot elkaar verhouden. Vervolgens werd gekeken hoe de twee belangrijke populaties, namelijk studenten en praktijkmensen met elkaar verschillen en als laatste wordt gekeken hoe deze verschillen van invloed zijn om het model te kunnen begrijpen. Een uitgebreide set aan onafhankelijke variabelen zijn verzameld door Mendling et al. (2018) op basis van literatuuronderzoek:

- Positie: het vastleggen van de positie van de participant in relatie tot het onderzoek: student, *practitioner* of anders.
- Theorie: niveau van theoretisch kennis van modellering.
- Modelleerjaren ervaring: aantal jaren ervaring in modellering.
- Modellen gelezen: aantal modellen gelezen/geanalyseerd in de afgelopen 12 maanden.
- Modellen gecreëerd: aantal modellen gecreëerd of aangepast in de afgelopen 12 maanden.
- Training: aantal dagen modelleertrainingen gevolgd in de afgelopen 12 maanden.
- Zelfstudie: aantal dagen zelfstudie gevolgd in de afgelopen 12 maanden.
- FAM1, FAM2, FAM3: niveau van bekendheid met een specifieke modelleringsgrammatica.
- Maanden BPNM: aantal maanden werkervaring met een specifieke modelleringsgrammatica.

Afhankelijke variabelen van het begrip van het conceptueel model zijn vastgelegd in:

- *Performance*: mate van nauwkeurigheid van het begrip van het model. Dit wordt berekend door de som van de juiste antwoorden op begripvragen die betrekking hebben op een bepaald model.
- *Voltooitijd*: deze variabele legt de tijd vast voor het voltooien van een specifieke begrijpelijke taak.

De studie is uitgevoerd als een online semi-experiment dat MVC's als *within-subject design* (i.e. goed vs fout) als *between-subject design* kenmerkte. De motivatie hiervoor is dat een willekeurige toewijzing van participanten aan groepen niet haalbaar was.

Als materiaal zijn er vijf BPMN modellen ontworpen met een goede en slechte *lay-out*. De uitdaging was om het proces tussen de verschillende activiteiten goed te begrijpen. De participanten waren willekeurig toegewezen aan een van de twee groepen:

- A: M1 (goed) - M2 (slecht) - M3 (goed) - M4 (slecht) - M5 (goed).
- B: M1 (slecht) - M2 (goed) - M3 (slecht) - M4 (goed) - M5 (slecht).

De participanten moesten per model zes vragen beantwoorden, in totaal 30 vragen. De vragen waren gericht op exclusiviteit, gelijktijdigheid en optionele activiteiten. Bij het beantwoorden van de begripstaken mochten de participanten bij het beantwoorden van de vragen het model erbij houden i.p.v. deze te moeten onthouden in het geheugen. In totaal zijn er 2199 personen die begonnen zijn aan de taken en hebben 778 personen dit compleet afgerond.

Het experiment bestond uit drie taken:

- Participanten beoordelen verschillende MVC modellen.
- Participanten leggen een theoretische kennistoets af.
- Participanten beantwoorden begripsvragen over de vijf procesmodellen (A of B groep).

H5 Artikel: Recker et al. (2011)

Recker et al. (2011) beschrijven in dit artikel hoe het onderzoek is verlopen. In het artikel worden twee belangrijke inzichten besproken die samenhangen met het gebruik van de grammatica, namelijk het waargenomen nut en het waargenomen gebruikersgemak. De onderzoeksvraag is als volgt beschreven: Hoe worden de waarnemingen van gebruikers van ontologische tekortkomingen die bestaan in een modelleringsgrammatica geassocieerd met hun overtuigingen over het nut en het gebruiksgemak van de grammatica?

In de studie zijn zes concepten gemeten: *perceived usefulness*, *perceived ease of use (PEOU)*, *perceived construct deficit (PCD)*, *perceived construct redundancy (PCR)*, *perceived construct overload (PCO)* en *perceived construct excess (PCE)*. De concepten zijn met behulp van 7-punts likert-schalen tussen *strongly agree* (i.e. gecodeerd als 1) en *strongly disagree* (i.e. gecodeerd als 7). PU en PEOU zijn de afhankelijke variabelen en PCD, PCR, PCO en PCE zijn de onafhankelijke variabelen.

Het verzamelen van empirische gegevens is via een *survey uitgevoerd* bij gebruikers van de BPMN in enkele maanden in 2007. Recker et al. (2011) maakten gebruik van de *survey methode*, omdat duidelijk geïdentificeerde onafhankelijke en afhankelijke variabelen bestaan en er een specifiek model aanwezig was dat de relatie tussen de variabelen onderbouwt. In totaal hebben er 590 participanten meegedaan, waarvan 60 onvolledige reacties en twee ongeldig. In totaal bleven er 528 bruikbare reacties over. De modelleerervaring van de respondenten waren onderverdeeld in vier categorieën: geen ervaring, weinig ervaring, ervaring en uitgebreide ervaring.

Voor het verzamelen van data over de BPMN grammatica werd gebruik gemaakt van een web gebaseerde *survey*. Het voordeel hiervan was lage kosten, geen geografische beperkingen en snelle response. De uitnodiging verliep via advertenties op online forums en blogs. Participanten werden geïnformeerd over het type onderzoek en werden gemotiveerd om deel te nemen door een samenvatting van de resultaten beschikbaar te maken en om gratis een modelleringsseminar bij te wonen en de kans op het winnen van een gratis lesboek.

Om rekening te houden met mogelijke hiaten in de perceptie kregen participanten 12 tekortkomingen ter overweging. Wanneer de participanten een situatie tegen kwamen waarin de tekortkomingen in gebruik zich zouden voordoen en de vraag met ja werd beantwoord werd gevraagd om de meetschalen in te vullen voor het waargenomen tekort. Met deze extra vraag (i.e. per tekortkoming) konden twee groepen worden geïdentificeerd. Een groep waar de tekortkoming in BPMN zich

voordeden en een groep waarin zich dit niet voordeed. Hierdoor ontstonden twee groepen die kruislings verhoord konden worden over het nut en het gebruiksgemak van de modelleringsgrammatica.

H6 Artikel: Shanks et al. (2008)

Shanks et al. (2008) hebben onderzoek gedaan naar eerdere theorieën over alternatieve weergaven van conceptuele modellering. Het model is uitgebreid met *part-whole* relaties en is empirisch geëvalueerd in het artikel. De aanleiding hiervoor was om te beoordelen of *entity-based* (i.e. *class-based*) of *relationship-based* (i.e. *association-based*) van *part-whole* relaties op basis van entiteiten of op basis van relaties in een conceptueel model eindgebruikers in staat te stellen om het model beter te begrijpen.

De propositie is als volgt omschreven “Conceptuele modellen die gebruik maken van een *entity class* om een compositie weer te geven, stellen hun gebruikers in staat om de semantiek van de compositie beter te begrijpen dan conceptuele diagrammen die gebruik maken van een *relationship class* voor het weergeven van de compositie” (Shanks et al., 2008).

Het artikel beschrijft dat het experiment werd geëvalueerd met probleemoplossende taken. Er zijn geen terugroep- en begripstaken uitgevoerd. Als motivatie hiervoor werd aangegeven dat probleemoplossende taken een betere indicator was om details te meten wat iemand zijn begrip is van het domein. De probleemoplossende eigenschappen zijn op drie manieren gemeten: (1) nauwkeurigheid van de oplossing, (2) de tijd die nodig is om een oplossing te vinden en (3) de genormaliseerde nauwkeurigheid (i.e. totale nauwkeurigheidsscore van de oplossing gedeeld door de totale tijd die nodig is). Elke probleem werd nauwkeuring geëvalueerd. De genormaliseerde nauwkeurigheid werd gebruikt om rekening te houden met het feit dat probleemoplossers de nauwkeurigheid van de oplossing kunnen afwegen tegen de tijd die nodig is om een taak uit te voeren. De probleemoplossende *performance* is gekenmerkt als de afhankelijke variabelen.

De participanten waren 30 personen die werkzaam zijn in een niet nader benoemde sector. De onderzoekers kende deze participanten en waren bereid om aan het experiment deel te nemen. Negen participanten hebben een informatie/technologie-achtergrond en werden gekozen, omdat ze in hun oorspronkelijke rol geen technische informatierol vervullen. Vier participanten waren bijna klaar met hun bachelor studie en hebben vijf jaar werkervaring. De overige participanten hadden minstens een bachelor. Er waren 14 participanten zonder modelleerervaring. De andere hadden wel ervaring met een of twee modelleertechnieken, zoals UML, flowchart, financiële modellen, data flow diagrammen en *entity relationship digrams*.

Voor het experiment zijn vijf verschillende materialen gebruik:

1. Samenvatting van de UML symbolen.
2. Twee UML diagrammen.
3. 11 probleemoplossende vragen, participanten konden hierop antwoorden met mogelijk, niet mogelijk of onzeker.
4. Drie vragen om de mening te meten over het gebruiksgemak van UML dat de participanten gebruikten bij het oplossen van het probleem.
5. Een set aan persoonlijke achtergrondvragen om informatie te verkrijgen over opleiding, sector, aantal jaar in dienst, aantal oorspronkelijke functies en de mate waarin zij eerder betrokken zijn geweest bij een vorm van modellering.

De participanten werden willekeurig toegewezen aan een van de twee groepen. Per groep zaten er 15 participanten in. Voor de start van het experiment werd een toestemmingsformulier ingevuld en achtergrond informatie vastgelegd. Na uitleg van het doel, aard van het experiment en de bevestiging

van de participanten dat ze vertrouwd waren met de UML symbolen, konden ze starten. Participanten kregen vervolgens de mogelijkheid om duidelijke of onduidelijke UML diagram door te nemen. Het diagram kon gedurende heel het experiment geraadpleegd worden, net zoals de samenvatting van de UML symbolen. Bij het uitvoeren van de probleemoplossende taken werd gevraagd om hardop te lezen en dit werd vervolgens opgenomen. De tijd tussen het beantwoorden van de vraag is ook genoteerd. Daarnaast werden er aantekeningen gemaakt op basis van de reacties van de participanten. Het experiment duurde gemiddeld 60 minuten per participant.

Bijlage D: Survey vragen

H1 Datasets en export van de vragen

In tabel 24 zijn de volgende bronnen opgenomen naar bestanden die onderdeel zijn van de *survey* over het evalueren van de PGA-notatie en maken onderdeel uit van dit onderzoek.

Tabel 24: Overzicht van externe bijlages

Nummer	Naam	Toepassing
1	Bijlage D Survey_archive.lsa	De complete export van de vragen en <i>timers</i> die gebruikt zijn in het onderzoek.
2	Bijlage D SPSS dataset persoonlijke resultaten.sav	SPSS dataset persoonlijke resultaten
3	Bijlage D SPSS dataset groepsresultaten.sav	SPSS dataset groepsresultaten
4	Bijlage D SPSS dataset persoonlijke resultaten geaggregeerd.sav	SPSS dataset persoonlijke resultaten geaggregeerd
5	Bijlage D SPSS dataset groepsresultaten geaggregeerd.sav	SPSS dataset groepsresultaten geaggregeerd

H2 Codelijst van de vragen

In deze bijlage is een lijst toegevoegd met de naam van het element, de *surveycode* en een korte omschrijving van de vraag. Hiermee is het mogelijk om te achterhalen welke vraag er bij het PGA-element gesteld is. De vragen zijn overgenomen uit de *survey*.

Tabel 25: Elementenlijst van de vragen

Element	Surveycode	Vraag
Declaration	Declaration	Declaration of consent for participation in scientific research [yes/no]
GenderCode	GenderCode	What is your gender?
YearOfBirth	YearOfBirth	Please select your year of birth
DegreeCode	DegreeCode	What is your highest degree?
JobSector	JobSectorCode	In which sector are you working?
ModelingExperience [SQ001]	ModelingExperience[SQ001]Code	Which conceptual modeling languages do you know? [None]
ModelingExperience [SQ002]	ModelingExperience[SQ002]Code	Which conceptual modeling languages do you know? [Requirements Engineering: e.g., Entity-Relationship (ER) models, UML class diagrams]
ModelingExperience [SQ003]	ModelingExperience[SQ003]Code	Which conceptual modeling languages do you know? [Business Process Management: e.g., BPMN, Event-driven Process Chains, UML activity diagrams, petri-nets]
ModelingExperience [SQ004].	ModelingExperience[SQ004]Code	Which conceptual modeling languages do you know? [Enterprise Architecture: e.g., ArchiMate]
ModelingExpertise	ModelingExpertiseCode	Please rate your level of modeling expertise.
ModelNumberRead	ModelNumberReadCode	How many conceptual models have you read by average during the last five years?
	ModelNumbersMade	How many conceptual models have you made by average during the last five years?
Internal goal	AV: Notation Association - Q1	This part of the research revolves around notation association. In this question, you are asked to record intuitive associations regarding achieving strategic fit, which occur to you after seeing a notation. You can name up to three terms. Please fill in at least one answer
Value proposition	AV: Notation Association - Q2	This part of the research revolves around notation association. In this question, you are asked to record intuitive associations regarding achieving strategic fit, which occur to you after seeing a notation. You can name up to three terms. Please fill in at least one answer
Competence	AV: Notation Association - Q3	This part of the research revolves around notation association. In this question, you are asked to record intuitive associations regarding achieving strategic fit, which occur to you after seeing a notation. You can name up to three terms. Please fill in at least one answer

Element	Surveycode	Vraag
Value stream	AV: Notation Association – Q4	This part of the research revolves around notation association. In this question, you are asked to record intuitive associations regarding achieving strategic fit, which occur to you after seeing a notation. You can name up to three terms. Please fill in at least one answer
Customer goal	AV: Notation Association – Q5	This part of the research revolves around notation association. In this question, you are asked to record intuitive associations regarding achieving strategic fit, which occur to you after seeing a notation. You can name up to three terms. Please fill in at least one answer
Customer goal	CQQ1Code	Which element is a customer goal?
Value proposition	CQQ2Code	Which type element is supported by 'partner relation development'?
Value stream	CQQ3Code	Which path is most important?
Competence	CQQ4Code	Which elements are competences?
Internal goal	CQQ5Code	Which type of element is increase innovation level?
Value Stream	CQQ6Code	How many value streams have a medium importance?
Internal goal	RQQ1Code	Record the concept name of the shown notation. Concept: internal goal
Value proposition	RQQ2Code	Record the concept name of the shown notation. Concept: value proposition
Competence	RQQ3Code	Record the concept name of the shown notation. Concept: Competence
Value stream	RQQ4Code	Record the concept name of the shown notation. Concept: value stream
Customer goal	RQQ5Code	Record the concept name of the shown notation. Concept: customer goal.
Customer goal	CustomerGoalCode	Compare the intuitiveness of alternative A and B relatively to each other by switching the slider bar on a scale from 0 to 100: * 0 indicates alternative A is more intuitive * 100 indicates alternative B is more intuitive * with 50 indicating no preference for either.
Value proposition	ValuePropositionCode	Compare the intuitiveness of alternative A and B relatively to each other by switching the slider bar on a scale from 0 to 100: * 0 indicates alternative A is more intuitive * 100 indicates alternative B is more intuitive * with 50 indicating no preference for either.
Competence	CompetenceCode	Compare the intuitiveness of alternative A and B relatively to each other by switching the slider bar on a scale from 0 to 100: * 0 indicates alternative A is more intuitive * 100 indicates alternative B is more intuitive * with 50 indicating no preference for either.
Internal goal	InternalGoalCode	Compare the intuitiveness of alternative A and B relatively to each other by switching the slider bar on a scale from 0 to 100: * 0 indicates alternative A is more intuitive * 100 indicates alternative B is more intuitive * with 50 indicating no preference for either.
Value stream	ValueStreamCode	Compare the intuitiveness of alternative A and B relatively to each other by switching the slider bar on a scale from 0 to 100: * 0 indicates alternative A is more intuitive * 100 indicates alternative B is more intuitive * with 50 indicating no preference for either.

Bijlage E: Persoonlijke resultaten van de notatie-associatievragen

H1 Normale verdeling

H1.1 Resultaten van de vragen

Voor het vaststellen of een variabele normaal verdeeld is, is de *Shapiro-Wilk* test toegepast. In tabel 26 zijn de persoonlijke resultaten weergegeven voor de notatie-associatievragen.

Tabel 26: Notatie-associatievragen: Persoonlijke resultaten Shapiro-Wilk test

	Kolmogorov-Smirnov ^a		Shapiro-Wilk			
	Statistic	df	Sig.	Statistic	df	Sig.
AV: Notation Association - Q1	.247	14	.021	.900	14	.113
AV: Notation Association - Q2	.510	14	.000	.428	14	.000
AV: Notation Association - Q3	.474	14	.000	.552	14	.000
AV: Notation Association - Q4	.392	14	.000	.693	14	.000
AV: Notation Association - Q5	.245	14	.022	.884	14	.067

a. Lilliefors Significance Correction

H1.2 Resultaten van de tijden

Voor het vaststellen of een variabele normaal verdeeld is, is de *Shapiro-Wilk* toegepast. In tabel 27 zijn de persoonlijke resultaten weergegeven voor de notatie-associatietijden.

Tabel 27: Notatie-associatietijden: persoonlijke resultaten Shapiro-Wilk test

	Kolmogorov-Smirnov ^a		Shapiro-Wilk			
	Statistic	df	Sig.	Statistic	df	Sig.
Group time: Notation association - Q1	.169	14	.200*	.917	14	.196
Group time: Notation association - Q2	.266	14	.008	.707	14	.000
Group time: Notation association - Q3	.201	14	.132	.846	14	.019
Group time: Notation association - Q4	.219	14	.068	.821	14	.009
Group time: Notation association - Q5	.363	14	.000	.483	14	.000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

H1.3 Samenvatting van de notatie-associatievragen

In tabel 28 zijn de persoonlijke resultaten van de *Shapiro-Wilk test* samengevat (tabel 26 en tabel 27) van de notatie-associatievragen. Als sig. < is dan 0.05 dan is de nulhypothese weerlegd en is de conclusie dat de verdeling afwijkend is van de normale verdeling.

Tabel 28: Notatie-associatieresultaten: persoonlijke resultaten Shapiro-Wilk test

Notatie-associatievragen en notatie-associatievragentijden	Sig.	Normaal verdeeld
AV: Notation Association - Q1	.113	Ja
AV: Notation Association - Q2	.000	Nee
AV: Notation Association - Q3	.000	Nee
AV: Notation Association - Q4	.000	Nee
AV: Notation Association - Q5	.067	Ja
Group time: Notation association - Q1	.196	Ja
Group time: Notation association - Q2	.000	Nee
Group time: Notation association - Q3	.019	Nee
Group time: Notation association - Q4	.009	Nee
Group time: Notation association - Q5	.000	Nee

H2 Mann-Whitney U test

H2.1 Resultaten van de vragen

Voor het vaststellen of de nieuwe versie beter is dan de oorspronkelijke versie is de *Mann-Whitney U Test* toegepast. In tabel 29 en tabel 30 zijn de persoonlijke resultaten weergegeven voor de notatie-associatievragen.

Tabel 29: Notatie-associatievragen: Persoonlijke resultaten Mann-Whitney U test (Ranks)

	Ranks			
	VersionCode	N	Mean Rank	Sum of Ranks
AV: Notation Association - Q2	0	7	6.50	45.50
	1	7	8.50	59.50
	Total	14		
AV: Notation Association - Q3	0	7	8.07	56.50
	1	7	6.93	48.50
	Total	14		
AV: Notation Association - Q4	0	7	8.43	59.00
	1	7	6.57	46.00
	Total	14		

Tabel 30: Notatie-associatievragen: persoonlijke resultaten Mann-Whitney U test (Test Statistics)

	Test Statistics ^a		
	AV: Notation Association - Q2	AV: Notation Association - Q3	AV: Notation Association - Q4
Mann-Whitney U	17.500	20.500	18.000
Wilcoxon W	45.500	48.500	46.000
Z	-1.472	-.713	-.971
Asymp. Sig. (2-tailed)	.141	.476	.332
Exact Sig. [2*(1-tailed Sig.)]	.383 ^b	.620 ^b	.456 ^b

a. Grouping Variable: VersionCode

b. Not corrected for ties.

H2.2 Resultaten van de notatie-associatietijden

Voor het vaststellen of de nieuwe notatie leidt tot een snellere associatie is de *Mann-Whitney U Test* toegepast. In tabel 31 en tabel 32 zijn de persoonlijke resultaten weergegeven voor de notatie-associatietijden.

Tabel 31: Notatie-associatietijden: persoonlijke resultaten Mann-Whitney U test (Ranks)

		Ranks		
	VersionCode	N	Mean Rank	Sum of Ranks
Group time: Notation association - Q2	0	7	7.00	49.00
	1	7	8.00	56.00
	Total	14		
Group time: Notation association - Q3	0	7	7.14	50.00
	1	7	7.86	55.00
	Total	14		
Group time: Notation association - Q4	0	7	8.29	58.00
	1	7	6.71	47.00
	Total	14		
Group time: Notation association - Q5	0	7	6.57	46.00
	1	7	8.43	59.00
	Total	14		

Tabel 32: Notatie-associatietijden: persoonlijke resultaten Mann-Whitney U test (Test Statistics)

Test Statistics ^a				
	Group time: Notation association - Q2	Group time: Notation association - Q3	Group time: Notation association - Q4	Group time: Notation association - Q5
Mann-Whitney U	21.000	22.000	19.000	18.000
Wilcoxon W	49.000	50.000	47.000	46.000
Z	-.447	-.319	-.703	-.831
Asymp. Sig. (2-tailed)	.655	.749	.482	.406
Exact Sig. [2*(1-tailed Sig.)]	.710 ^b	.805 ^b	.535 ^b	.456 ^b

a. Grouping Variable:

VersionCode

b. Not corrected for ties.

H3 Independent Samples Test

H3.1 Resultaten van de notatie-associatievragen

Voor het vaststellen of de nieuwe versie beter is dan de oorspronkelijke versie is de *Independent Samples Test* toegepast. In tabel 33 en tabel 34 zijn de persoonlijke resultaten weergegeven voor de notatie-associatievragen.

Tabel 33: Notatie-associatievragen: persoonlijke resultaten Independent Samples Test (Group Statistics)

Group Statistics					
	VersionCode	N	Mean	Std. Deviation	Std. Error Mean
AV: Notation Association - Q1	0	7	.142857142857143	.325300024316178	.122951852260553
	1	7	-.047619047619048	.621484823823870	.234899183919821
AV: Notation Association - Q5	0	7	.404761904761905	.089087080637475	.033671751485074
	1	7	.404761904761905	.358199495123707	.135386683406603

Tabel 34: Notatie-associatievragen: persoonlijke resultaten Independent Samples Test (Independent Samples Test)

Independent Samples Test										
		Levene's Test for Equality of Variances				t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
AV: Notation Association - Q1	Equal variances assumed	1.419	.257	.718	12	.486	.190476190476190	.265131636325239	-387196020299906	.768148401252287
	Equal variances not assumed			.718	9.058	.491	.190476190476190	.265131636325239	-408707077089781	.789659458042162
AV: Notation Association - Q5	Equal variances assumed	6.035	.030	.000	12	1.000	.000000000000000	.139511078025769	-303968526683252	.303968526683252
	Equal variances not assumed			.000	6.739	1.000	.000000000000000	.139511078025769	-332494500566302	.332494500566302

H3.2 Resultaten van de notatie-associatietijden

Voor het vaststellen of de nieuwe notatie leidt tot een snellere associatie is de *Independent Samples Test* toegepast. In tabel 35 en tabel 36 zijn de persoonlijke resultaten weergegeven voor de notatie-associatietijden.

Tabel 35: Notatie-associatietijden: persoonlijke resultaten Independent Samples Test (Group Statistics)

Group Statistics					
	VersionCode	N	Mean	Std. Deviation	Std. Error Mean
Group time: Notation association - Q1	0	7	94.368571428571	57.213331321925	21.624606622193
			430	146	753
	1	7	68.081428571428	38.128320370634	14.411150515613
			570	135	716

Tabel 36: Notatie-associatietijden: persoonlijke resultaten Independent Samples Test (Independent Samples Test)

		Levene's Test for Equality of Variances		t-test for Equality of Means			95% Confidence Interval of the Difference			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Group time:	Equal variances assumed	2.207	.163	1.012	12	.332	26.287142857142850	25.986628691469370	-30.332857135629040	82.907142849914750
Notation association - Q1	Equal variances not assumed			1.012	10.451	.335	26.287142857142850	25.986628691469370	-31.277454474704157	83.851740188989870

Bijlage F: Groepsresultaten van de notatie-associatievragen

H1 Normale verdeling

H1.1 Resultaten van de notatie-associatievragen.

Voor het vaststellen of een variabele normaal verdeeld is, is de *Shapiro-Wilk* toegepast. In tabel 37 zijn de groepsresultaten weergegeven voor de notatie-associatievragen.

Tabel 37: Notatie-associatievragen: groepsresultaten Shapiro-Wilk test

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
AV: Notation Association - Q1	.351	44	.000	.764	44	.000
AV: Notation Association - Q2	.537	44	.000	.276	44	.000
AV: Notation Association - Q3	.535	44	.000	.284	44	.000
AV: Notation Association - Q4	.387	44	.000	.719	44	.000
AV: Notation Association - Q5	.391	44	.000	.666	44	.000

a. Lilliefors Significance Correction

H1.2 Resultaten van de notatie-associatietijden

Voor het vaststellen of een variabele normaal verdeeld is, is de *Shapiro-Wilk* toegepast. In tabel 38 zijn de groepsresultaten weergegeven voor de notatie-associatietijden.

Tabel 38: Notatie-associatietijden: groepsresultaten Shapiro-Wilk test

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group time: Notation association - Q1	.191	44	.000	.705	44	.000
Group time: Notation association - Q2	.290	44	.000	.515	44	.000
Group time: Notation association - Q3	.204	44	.000	.685	44	.000
Group time: Notation association - Q4	.249	44	.000	.638	44	.000
Group time: Notation association - Q5	.319	44	.000	.357	44	.000

a. Lilliefors Significance Correction

H1.3 Samenvatting van de notatie-associatievragen

In Tabel 39 zijn de persoonlijke resultaten van de *Shapiro-Wilk Test* samengevat (zie tabel 37 en tabel 38). Als sig. < is dan 0.05 dan is de nulhypothese weerlegd en is de conclusie dat de verdeling afwijkend is van de normale verdeling.

Tabel 39: Notatie-associatievragen: persoonlijke resultaten Shapiro-Wilk Test

Notatie-associatievragen en notatie-associatietijden	Sig.	Normaal verdeeld
AV: Notation Association - Q1	.000	Nee
AV: Notation Association - Q2	.000	Nee
AV: Notation Association - Q3	.000	Nee
AV: Notation Association - Q4	.000	Nee
AV: Notation Association - Q5	.000	Nee

Notatie-associatievragen en notatie-associatietijden	Sig.	Normaal verdeeld
Group time: Notation association - Q1	.000	Nee
Group time: Notation association - Q2	.000	Nee
Group time: Notation association - Q3	.000	Nee
Group time: Notation association - Q4	.000	Nee
Group time: Notation association - Q5	.000	Nee

H2 Mann-Whitney U test

H2.1 Resultaten van de notatie-associatievragen

Voor het vaststellen of de nieuwe versie beter is dan de oorspronkelijke versie is de Mann-Whitney U Test toegepast. In tabel 40 en tabel 41 zijn de groepsresultaten weergegeven voor de notatie-associatievragen.

Tabel 40: Notatie-associatievragen: groepsresultaten Mann-Whitney U test (Ranks)

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
AV: Notation Association - Q1	0	24	23.13	555.00
	1	20	21.75	435.00
	Total	44		
AV: Notation Association - Q2	0	24	21.00	504.00
	1	20	24.30	486.00
	Total	44		
AV: Notation Association - Q3	0	24	22.85	548.50
	1	20	22.08	441.50
	Total	44		
AV: Notation Association - Q4	0	24	24.23	581.50
	1	20	20.43	408.50
	Total	44		
AV: Notation Association - Q5	0	24	23.10	554.50
	1	20	21.78	435.50
	Total	44		

Tabel 41: Notatie-associatievragen: groepsresultaten Mann-Whitney U test (Test Statistics)

	Test Statistics ^a				
	AV: Notation Association - Q1	AV: Notation Association - Q2	AV: Notation Association - Q3	AV: Notation Association - Q4	AV: Notation Association - Q5
Mann-Whitney U	225.000	204.000	231.500	198.500	225.500
Wilcoxon W	435.000	504.000	441.500	408.500	435.500
Z	-.429	-1.943	-.459	-1.214	-.406
Asymp. Sig. (2-tailed)	.668	.052	.647	.225	.684

a. Grouping Variable: VersionCode

H2.2 Resultaten van de notatie-associatietijden

Voor het vaststellen of de nieuwe notatie leidt tot een snellere associatie is de *Mann-Whitney U Test* toegepast. In tabel 42 en tabel 43 zijn de groepsresultaten weergegeven voor de notatie-associatietijden.

Tabel 42: Notatie-associatietijden: groepsresultaten Mann-Whitney U test (Ranks)

		anks		
	VersionCode	N	Mean Rank	Sum of Ranks
Group time: Notation association - Q1	0	24	21.92	526.00
	1	20	23.20	464.00
	Total	44		
Group time: Notation association - Q2	0	24	20.58	494.00
	1	20	24.80	496.00
	Total	44		
Group time: Notation association - Q3	0	24	22.75	546.00
	1	20	22.20	444.00
	Total	44		
Group time: Notation association - Q4	0	24	24.88	597.00
	1	20	19.65	393.00
	Total	44		
Group time: Notation association - Q5	0	24	20.17	484.00
	1	20	25.30	506.00
	Total	44		

Tabel 43: Notatie-associatietijden: groepsresultaten Mann-Whitney U test (Test Statistics)

Test Statistics ^a					
	Group time: Notation association - Q1	Group time: Notation association - Q2	Group time: Notation association - Q3	Group time: Notation association - Q4	Group time: Notation association - Q5
Mann-Whitney U	226.000	194.000	234.000	183.000	184.000
Wilcoxon W	526.000	494.000	444.000	393.000	484.000
Z	-.330	-1.084	-.141	-1.344	-1.320
Asymp. Sig. (2-tailed)	.741	.278	.888	.888	.187

a. Grouping Variable:

VersionCode

Bijlage G: Persoonlijke resultaten van de begripsvragen

H1 Normaalverdeling

H1.1 Resultaten van de begripsvragentijden

Voor het vaststellen of een variabele normaal verdeeld is, is de *Shapiro-Wilk* toegepast. In tabel 44 zijn de persoonlijke resultaten weergegeven voor de tijden van de begripsvragen.

Tabel 44: Begripsvragentijden: persoonlijke resultaten Shapiro-Wilk test

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group time: Comprehension questions - Q1	.119	14	.200*	.957	14	.678
Group time: Comprehension questions - Q2	.245	14	.022	.687	14	.000
Group time: Comprehension questions - Q3	.361	14	.000	.485	14	.000
Group time: Comprehension questions - Q4	.206	14	.110	.882	14	.061
Group time: Comprehension questions - Q5	.394	14	.000	.533	14	.000
Group time: Comprehension questions - Q6	.127	14	.200*	.970	14	.876

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

H1.2 Samenvatting van de begripsvragentijden

In tabel 45 zijn de persoonlijke resultaten van de *Shapiro-Wilk Test* uit tabel 44 samengevat. Als sig. < 0.05 is dan de nulhypothese weerlegd en is de conclusie dat de verdeling afwijkend is van de normale verdeling.

Tabel 45: Begripsvragentijden: persoonlijke resultaten Shapiro-Wilk test

Begripsvragentijden	Sig.	Normaal verdeeld
Group time: Comprehension questions - Q1	.678	Ja
Group time: Comprehension questions - Q2	.000	Nee
Group time: Comprehension questions - Q3	.000	Nee
Group time: Comprehension questions - Q4	.061	Ja
Group time: Comprehension questions - Q5	.000	Nee
Group time: Comprehension questions - Q6	.876	Ja

H2 Mann-Whitney U test

H2.1 Resultaten van de begripsvragen

Voor het vaststellen of de nieuwe versie beter is dan de oorspronkelijke versie is de *Mann-Whitney U Test* toegepast. In tabel 46 en tabel 47 zijn de persoonlijke resultaten weergegeven voor de begripsvragen.

Tabel 46: Begripsvragen: persoonlijke resultaten Mann-Whitney U test (Ranks)

	VersionCode	Ranks		
		N	Mean Rank	Sum of Ranks
CQQ1Code	0	7	8.00	56.00
	1	7	7.00	49.00
	Total	14		
CQQ2Code	0	7	8.00	56.00
	1	7	7.00	49.00
	Total	14		
CQQ3Code	0	7	7.50	52.50
	1	7	7.50	52.50
	Total	14		
CQQ4Code	0	7	8.00	56.00
	1	7	7.00	49.00
	Total	14		
CQQ5Code	0	7	7.00	49.00
	1	7	8.00	56.00
	Total	14		
CQQ6Code	0	7	8.00	56.00
	1	7	7.00	49.00
	Total	14		

Tabel 47: Begripsvragen: persoonlijke resultaten Mann-Whitney U test (Test statistics)

	Test Statistics ^a					
	CQQ1Code	CQQ2Code	CQQ3Code	CQQ4Code	CQQ5Code	CQQ6Code
Mann-Whitney U	21.000	21.000	24.500	21.000	21.000	21.000
Wilcoxon W	49.000	49.000	52.500	49.000	49.000	49.000
Z	-1.000	-.515	.000	-.537	-1.000	-1.000
Asymp. Sig. (2-tailed)	.317	.606	1.000	.591	.317	.317
Exact Sig. [2*(1-tailed Sig.)]	.710 ^b	.710 ^b	1.000 ^b	.710 ^b	.710 ^b	.710 ^b

a. Grouping Variable: VersionCode

b. Not corrected for ties.

H2.2 Resultaten van de begripsvragentijden

Voor het vaststellen of de nieuwe notatie leidt tot een snellere associatie is de *Mann-Whitney U Test* toegepast. In tabel 48 en tabel 49 zijn de persoonlijke resultaten weergegeven voor de begripsvragentijden.

Tabel 48: Begripsvragentijden: persoonlijke resultaten Mann-Whitney U test (Ranks)

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
Group time: Comprehension questions - Q2	0	7	7.57	53.00
	1	7	7.43	52.00
	Total	14		
Group time: Comprehension questions - Q3	0	7	6.57	46.00
	1	7	8.43	59.00
	Total	14		
Group time: Comprehension questions - Q5	0	7	6.00	42.00
	1	7	9.00	63.00
	Total	14		

Tabel 49: Begripsvragentijden: persoonlijke resultaten Mann-Whitney U test (Test statistics)

	Test Statistics ^a		
	Group time: Comprehension questions - Q2	Group time: Comprehension questions - Q3	Group time: Comprehension questions - Q5
Mann-Whitney U	24.000	18.000	14.000
Wilcoxon W	52.000	46.000	42.000
Z	-.064	-.831	-1.342
Asymp. Sig. (2-tailed)	.949	.406	.180
Exact Sig. [2*(1-tailed Sig.)]	1.000 ^b	.456 ^b	.209 ^b

a. Grouping Variable: VersionCode

b. Not corrected for ties.

H3 Independent Samples Test

H3.1 Resultaten van de begripsvragentijden

Voor het vaststellen of de nieuwe notatie leidt tot een snellere associatie is de *Independent Samples Test* toegepast. In tabel 50 en tabel 51 zijn de persoonlijke resultaten weergegeven voor de begripsvragentijden.

Tabel 50: Begripsvragentijden: persoonlijke resultaten Independent Samples Test (Group Statistics)

Group Statistics					
	VersionCode	N	Mean	Std. Deviation	Std. Error Mean
Group time: Comprehension questions - Q1	0	7	46.35571428571	25.20498684595	9.526589570435
			4285	4950	865
	1	7	42.36000000000	13.47879816600	5.094506845613
			0000	8717	224
Group time: Comprehension questions - Q4	0	7	93.19857142857	77.03926518959	29.11810526840
			1440	2640	2488

	1	7	92.11428571428 5730	52.31533390799 2620	19.77333761083 6187
Group time: Comprehension questions - Q6	0	7	39.06571428571 4290	16.16102086002 2898	6.108291732649 683
	1	7	39.19428571428 5720	13.15420573700 3986	4.971822439241 665

Tabel 51: Begripsvragentijden: persoonlijke resultaten Independent Samples Test (Independent Samples Test)

Independent Samples Test

		Levene's Test for Equality of Variances					t-test for Equality of Means			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Group time: Comprehension questions - Q1	Equal variances assumed	5.806	.033	.370	12	.718	3.995714285714286	10.803236035722696	-19.542514990841620	27.533943562270192
	Equal variances not assumed			.370	9.172	.720	3.995714285714286	10.803236035722696	-20.373112819569440	28.364541390998006
Group time: Comprehension questions - Q4	Equal variances assumed	1.171	.300	.031	12	.976	1.084285714285713	35.197285899538870	-75.604012373094890	77.772583801666330
	Equal variances not assumed			.031	10.563	.976	1.084285714285713	35.197285899538870	-76.776958391530810	78.945529820102240
Group time: Comprehension questions - Q6	Equal variances assumed	.034	.858	-	12	.987	-.128571428571426	7.875928279161981	-17.288745008748464	17.031602151605610
	Equal variances not assumed			-	11.525	.987	-.128571428571426	7.875928279161981	-17.367504847323530	17.110361990180678

Bijlage H: Groepsresultaten van de begripsvragen

H1 Normaleverdeling

H1.1 Resultaten van de begripsvragen

Voor het vaststellen of een variabele normaal verdeeld is, is de *Shapiro-Wilk* toegepast. In tabel 52 zijn de groepsresultaten weergegeven voor de begripsvragen.

Tabel 52: Begripsvragen: groepsresultaten Shapiro-Wilk test

	Kolmogorov-Smirnov ^a		Shapiro-Wilk			
	Statistic	df	Sig.	Statistic	df	Sig.
Group time: Comprehension questions - Q1	.211	44	.000	.724	44	.000
Group time: Comprehension questions - Q2	.197	44	.000	.757	44	.000
Group time: Comprehension questions - Q3	.334	44	.000	.380	44	.000
Group time: Comprehension questions - Q4	.199	44	.000	.881	44	.000
Group time: Comprehension questions - Q5	.265	44	.000	.610	44	.000
Group time: Comprehension questions - Q6	.213	44	.000	.658	44	.000

a. Lilliefors Significance Correction

H1.2 Samenvatting van de begripsvragentijden

In tabel 53 zijn de groepsresultaten van *Independent Samples Test* (zie tabel 52) van de begripsvragentijden weergegeven. Als sig. < is dan 0.05 dan is de nulhypothese weerlegd en is de conclusie dat de verdeling afwijkend is van de normale verdeling.

Tabel 53: Begripsvragentijden: groepsresultaten Shapiro-Wilk test

Begripsvragentijden	Sig.	Normaal verdeeld
Group time: Comprehension questions - Q1	.000	Nee
Group time: Comprehension questions - Q2	.000	Nee
Group time: Comprehension questions - Q3	.000	Nee
Group time: Comprehension questions - Q4	.000	Nee
Group time: Comprehension questions - Q5	.000	Nee
Group time: Comprehension questions - Q6	.000	Nee

H2 Mann-Whitney U test

H2.1 Resultaten van de begripsvragen

Voor het vaststellen of de nieuwe versie beter is dan de oorspronkelijke versie is de *Mann-Whitney U Test* toegepast. In tabel 54 en tabel 55 zijn de groepsresultaten weergegeven voor de begripsvragen.

Tabel 54: Begripsvragen: groepsresultaten Mann-Whitney U test (Ranks)

	VersionCode	Ranks		
		N	Mean Rank	Sum of Ranks
CQQ1Code	0	24	23.50	564.00
	1	20	21.30	426.00
	Total	44		
CQQ2Code	0	24	26.25	630.00
	1	20	18.00	360.00
	Total	44		
CQQ3Code	0	24	22.58	542.00
	1	20	22.40	448.00
	Total	44		
CQQ4Code	0	24	22.33	536.00
	1	20	22.70	454.00
	Total	44		
CQQ5Code	0	24	21.42	514.00
	1	20	23.80	476.00
	Total	44		
CQQ6Code	0	24	22.50	540.00
	1	20	22.50	450.00
	Total	44		

Tabel 55: Begripsvragen: groepsresultaten Mann-Whitney U test (Test Statistics)

	Test Statistics ^a					
	CQQ1Code	CQQ2Code	CQQ3Code	CQQ4Code	CQQ5Code	CQQ6Code
Mann-Whitney U	216.000	150.000	238.000	236.000	214.000	240.000
Wilcoxon W	426.000	360.000	448.000	536.000	514.000	450.000
Z	-1.568	-2.459	-.131	-.111	-.967	.000
Asymp. Sig. (2-tailed)	.117	.014	.896	.912	.333	1.000

a. Grouping Variable: VersionCode

H2.2 Resultaten van de begripsvragentijden

Voor het vaststellen of de nieuwe notatie leidt tot een snellere associatie is de *Mann-Whitney U Test* toegepast. In tabel 56 en tabel 57 zijn de groepsresultaten weergegeven voor de begripsvragentijden.

Tabel 56: Begripsvragentijden: groepsresultaten Mann-Whitney U Test (Ranks)

Ranks				
	VersionCode	N	Mean Rank	Sum of Ranks
Group time: Comprehension questions - Q1	0	24	22.13	531.00
	1	20	22.95	459.00
	Total	44		
Group time: Comprehension questions - Q2	0	24	22.67	544.00
	1	20	22.30	446.00
	Total	44		
Group time: Comprehension questions - Q3	0	24	20.69	496.50
	1	20	24.68	493.50
	Total	44		
Group time: Comprehension questions - Q4	0	24	20.00	480.00
	1	20	25.50	510.00
	Total	44		
Group time: Comprehension questions - Q5	0	24	21.13	507.00
	1	20	24.15	483.00
	Total	44		
Group time: Comprehension questions - Q6	0	24	27.17	652.00
	1	20	16.90	338.00
	Total	44		

Tabel 57: Begripsvragentijden: groepsresultaten Mann-Whitney U Test (Test Statistics)

	Test Statistics^a					
	Group time: Comprehension questions - Q1	Group time: Comprehension questions - Q2	Group time: Comprehension questions - Q3	Group time: Comprehension questions - Q4	Group time: Comprehension questions - Q5	Group time: Comprehension questions - Q6
Mann-Whitney U	231.000	236.000	196.500	180.000	207.000	128.000
Wilcoxon W	531.000	446.000	496.500	480.000	507.000	338.000
Z	-.212	-.094	-1.025	-1.414	-.778	-2.640
Asymp. Sig. (2-tailed)	.832	.925	.305	.157	.437	.008

a. Grouping Variable: VersionCode

Bijlage I: Persoonlijke resultaten van herinneringsvragen

H1 Normale verdeling

H1.1 Resultaten van de herinneringsvragentijden

Voor het vaststellen of een variabele normaal verdeeld is, is de *Shapiro-Wilk* toegepast. In tabel 58 zijn de persoonlijke resultaten weergegeven voor de herinneringsvragen.

Tabel 58: Herinneringsvragentijden: persoonlijke resultaten Shapiro-Wilk test

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group time: Recall questions - Q1	.123	14	.200*	.951	14	.578
Group time: Recall questions - Q2	.166	14	.200*	.933	14	.340
Group time: Recall questions - Q3	.272	14	.006	.794	14	.004
Group time: Recall questions - Q4	.252	14	.016	.788	14	.004
Group time: Recall questions - Q5	.174	14	.200*	.941	14	.435

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

H1.2 Samenvatting van de herinneringsvragentijden

Voor het vaststellen of een variabele normaal verdeeld is, is de *Shapiro-Wilk* toegepast. In tabel 59 zijn de persoonlijke resultaten weergegeven voor de herinneringsvragentijden. Als sig. < is dan 0.05 dan is de nulhypothese weerlegd en is de conclusie dat de verdeling afwijkend is van de normale verdeling.

Tabel 59: herinneringsvragen resultaten: persoonlijke resultaten Shapiro-Wilk test

Herinneringsvragentijden	Sig.	Normaal verdeeld
Group time: Recall questions - Q1	.578	Ja
Group time: Recall questions - Q2	.340	Ja
Group time: Recall questions - Q3	.004	Nee
Group time: Recall questions - Q4	.004	Nee
Group time: Recall questions - Q5	.435	ja

H2 Mann-Whitney U test

H2.1 resultaten van de herinneringsvragen

Voor het vaststellen of de nieuwe versie beter is dan de oorspronkelijke versie is de *Mann-Whitney U Test* toegepast. In tabel 60 en tabel 61 zijn de persoonlijke resultaten weergegeven voor de herinneringsvragen.

Tabel 60: Herinneringsvragen: persoonlijke resultaten Mann-Whitney U test (Ranks)

	VersionCode	Ranks		
		N	Mean Rank	Sum of Ranks
RQQ1Code	0	7	6.57	46.00
	1	7	8.43	59.00
	Total	14		
RQQ2Code	0	7	9.64	67.50
	1	7	5.36	37.50
	Total	14		
RQQ3Code	0	7	8.14	57.00
	1	7	6.86	48.00
	Total	14		
RQQ4Code	0	7	8.07	56.50
	1	7	6.93	48.50
	Total	14		
RQQ5Code	0	7	9.07	63.50
	1	7	5.93	41.50
	Total	14		

Tabel 61: Herinneringsvragen: persoonlijke resultaten Mann-Whitney U test (Test Statistics)

	Test Statistics ^a				
	RQQ1Code	RQQ2Code	RQQ3Code	RQQ4Code	RQQ5Code
Mann-Whitney U	18.000	9.500	20.000	20.500	13.500
Wilcoxon W	46.000	37.500	48.000	48.500	41.500
Z	-1.042	-2.137	-.680	-.713	-1.580
Asymp. Sig. (2-tailed)	.297	.033	.496	.476	.114
Exact Sig. [2*(1-tailed Sig.)]	.456 ^b	.053 ^b	.620 ^b	.620 ^b	.165 ^b

a. Grouping Variable: VersionCode

b. Not corrected for ties.

H2.2 Resultaten van de herinneringsvragentijden

Voor het vaststellen of de nieuwe notatie leidt tot een snellere associatie is de *Mann-Whitney U Test* toegepast. In tabel 62 en tabel 63 zijn de persoonlijke resultaten weergegeven voor de herinneringsvragentijden.

Tabel 62: herinneringsvragentijden: persoonlijke resultaten Mann-Whitney U test (Ranks)

		Ranks		
	VersionCode	N	Mean Rank	Sum of Ranks
Group time: Recall questions - Q3	0	7	6.14	43.00
	1	7	8.86	62.00
	Total	14		
Group time: Recall questions - Q4	0	7	7.00	49.00
	1	7	8.00	56.00
	Total	14		

Tabel 63: Herinneringsvragentijden: persoonlijke resultaten Mann-Whitney U test (Test Statistics)

Test Statistics ^a		
	Group time: Recall questions- Q3	Group time: Recall questions- Q4
Mann-Whitney U	15.000	21.000
Wilcoxon W	43.000	49.000
Z	-1.214	-.447
Asymp. Sig. (2-tailed)	.225	.655
Exact Sig. [2*(1-tailed Sig.)]	.259 ^b	.710 ^b

a. Grouping Variable: VersionCode

b. Not corrected for ties.

H3 Independent Samples Test

H3.1 Resultaten van de herinneringsvragentijden

Voor het vaststellen of de nieuwe notatie leidt tot een snellere associatie is de *Independent Samples Test* toegepast. In tabel 64 en tabel 65 zijn de persoonlijke resultaten weergegeven voor de herinneringsvragentijden.

Tabel 64: herinneringsvragentijden: persoonlijke resultaten Independent Samples Test (Group Statistics)

		Group Statistics			
	VersionCode	N	Mean	Std. Deviation	Std. Error Mean
Group time: Recall questions - Q1	0	7	17.002857142857	8.4986856126613	3.2122012288606
			142	56	51
	1	7	16.795714285714	4.7967727245960	1.8130096749969
			286	45	79
Group time: Recall questions - Q2	0	7	15.672857142857	5.4127279452715	2.0458188653768
			143	72	87
	1	7	14.734285714285	8.8686298399524	3.3520270037715
			715	62	39
	0	7	11.0214	5.54226	2.09478

Group time: Recall questions - 1 Q5	7	14.8029	5.10640	1.93004
--	---	---------	---------	---------

Tabel 65: Herinneringsvragentijden: persoonlijke resultaten Independent Samples Test (Independent Samples Test)

		Independent Samples Test									
		Levene's Test for Equality of Variances					t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
									Lower	Upper	
Group time: Recall questions - Q1	Equal variances assumed	2.486	.141	.056	12	.956	.207142857142856	3.688528272404392	-7.829469865362119	8.243755579647830	
	Equal variances not assumed			.056	9.471	.956	.207142857142856	3.688528272404392	-8.074116615939891	8.488402330225602	
Group time: Recall questions - Q2	Equal variances assumed	1.010	.335	.239	12	.815	.938571428571429	3.927016662040738	-7.617662856999893	9.494805714142750	
	Equal variances not assumed			.239	9.925	.816	.938571428571429	3.927016662040738	-7.820303066717424	9.697445923860280	
Group time: Recall questions - Q5	Equal variances assumed	.029	.867	-	12	.209	-3.78143	2.84836	-9.98746	2.42461	
	Equal variances not assumed			-	11.920	.209	-3.78143	2.84836	-9.99206	2.42921	

Bijlage J: Groepsresultaten van herinneringsvragen

H1 Normale verdeling

H1.1 Resultaten van de herinneringsvragentijden

Voor het vaststellen of een variabele normaal verdeeld is, is de *Shapiro-Wilk* toegepast. In tabel 66 zijn de groepsresultaten weergegeven voor de herinneringsvragentijden.

Tabel 66: Herinneringsvragentijden: groepsresultaten Shapiro-Wilk test

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Group time: Recall questions - Q1	.187	44	.001	.707	44	.000
Group time: Recall questions - Q2	.176	44	.001	.850	44	.000
Group time: Recall questions - Q3	.181	44	.001	.819	44	.000
Group time: Recall questions - Q4	.225	44	.000	.642	44	.000
Group time: Recall questions - Q5	.177	44	.001	.853	44	.000

a. Lilliefors Significance Correction

H1.2 Samenvatting van de herinneringsvragentijden

In tabel 67 zijn de resultaten van de *Shapiro-Wilk Test* weergegeven (zie tabel 66). Als sig. < is dan 0.05 dan is de nulhypothese weerlegd en is de conclusie dat de verdeling afwijkend is van de normale verdeling.

Tabel 67: Herinneringsvragentijden: groepsresultaten Shapiro-Wilk test

Herinneringsvragentijden	Sig.	Normaal verdeeld
Group time: Recall questions - Q1	.000	Nee
Group time: Recall questions - Q2	.000	Nee
Group time: Recall questions - Q3	.000	Nee
Group time: Recall questions - Q4	.000	Nee
Group time: Recall questions - Q5	.000	Nee

H2 Mann-Whitney U test

H2.1 Resultaten van de herinneringsvragen

Voor het vaststellen of de nieuwe versie beter is dan de oorspronkelijke versie is de *Mann-Whitney U Test* toegepast. In tabel 68 en tabel 69 zijn de groepsresultaten weergegeven voor de herinneringsvragen.

Tabel 68: Herinneringsvragen: groepsresultaten Mann-Whitney U test (Ranks)

	VersionCode	Ranks		
		N	Mean Rank	Sum of Ranks
RQQ1Code	0	24	22.17	532.00
	1	20	22.90	458.00
	Total	44		
RQQ2Code	0	24	25.67	616.00
	1	20	18.70	374.00
	Total	44		
RQQ3Code	0	24	23.63	567.00
	1	20	21.15	423.00
	Total	44		
RQQ4Code	0	24	25.58	614.00
	1	20	18.80	376.00
	Total	44		
RQQ5Code	0	24	26.19	628.50
	1	20	18.08	361.50
	Total	44		

Tabel 69: Herinneringsvragen: groepsresultaten Mann-Whitney U test (Test Statistics)

	Test Statistics ^a				
	RQQ1Code	RQQ2Code	RQQ3Code	RQQ4Code	RQQ5Code
Mann-Whitney U	232.000	164.000	213.000	166.000	151.500
Wilcoxon W	532.000	374.000	423.000	376.000	361.500
Z	-.204	-1.956	-.729	-2.032	-2.240
Asymp. Sig. (2-tailed)	.838	.051	.466	.042	.025

a. Grouping Variable: VersionCode

H2.2 Resultaten herinneringsvragentijden

Voor het vaststellen of de nieuwe notatie leidt tot een snellere associatie is de *Independent Samples Test* toegepast. In tabel 70 en tabel 71 zijn de groepsresultaten weergegeven voor de herinneringsvragentijden.

Tabel 70: Herinneringsvragentijden: groepsresultaten Mann-Whitney U test (ranks)

	VersionCode	Ranks		
		N	Mean Rank	Sum of Ranks
Group time: Recall questions - Q1	0	24	19.50	468.00
	1	20	26.10	522.00
	Total	44		
Group time: Recall questions - Q2	0	24	23.71	569.00

	1	20	21.05	421.00
	Total	44		
Group time: Recall questions - Q3	0	24	19.58	470.00
	1	20	26.00	520.00
	Total	44		
Group time: Recall questions - Q4	0	24	21.77	522.50
	1	20	23.38	467.50
	Total	44		
Group time: Recall questions - Q5	0	24	21.02	504.50
	1	20	24.28	485.50
	Total	44		

Tabel 71: Herinneringsvragentijden: groepsresultaten Mann-Whitney U test (Test Statistics)

	Test Statistics ^a				
	Group time: Recall questions - Q1	Group time: Recall questions - Q2	Group time: Recall questions - Q3	Group time: Recall questions - Q4	Group time: Recall questions - Q5
Mann-Whitney U	168.000	211.000	170.000	222.500	204.500
Wilcoxon W	468.000	421.000	470.000	522.500	504.500
Z	-1.697	-.684	-1.650	-.412	-.837
Asymp. Sig. (2-tailed)	.090	.494	.099	.680	.403

a. Grouping Variable: VersionCode

Bijlage K: Persoonlijke resultaten van de representatievoorkeursvragen

H1 Normale verdeling

H1.1 Resultaten van de Representatievoorkeursvragen

Voor het vaststellen of een variabele normaal verdeeld is, is de *Shapiro-Wilk* toegepast. In tabel 72 zijn de persoonlijke resultaten weergegeven voor representatievoorkeursvragen.

Tabel 72: Representatievoorkeursvragen: persoonlijke resultaten Shapiro-Wilk test

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
CustomerGoalCode	.273	14	.006	.873	14	.046
ValuePropositionCode	.229	14	.044	.869	14	.040
CompetenceCode	.277	14	.005	.828	14	.011
InternalGoalCode	.168	14	.200*	.941	14	.430
ValueStreamCode	.200	14	.132	.893	14	.089

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

H1.2 Samenvatting van de representatievoorkeursvragen

In tabel 73 zijn de persoonlijke resultaten van de *Shapiro-Wilk* test weergegeven. Als sig. < is dan 0.05 dan is de nulhypothese weerlegd en is de conclusie dat de verdeling afwijkend is van de normale verdeling.

Tabel 73: Representatievoorkeursvragen: persoonlijke resultaten Shapiro-Wilk test

Representatievoorkeursvragen	Sig.	Normaal verdeeld
CustomerGoalCode	.046	Nee
ValuePropositionCode	.040	Nee
CompetenceCode	.011	Nee
InternalGoalCode	.430	Ja
ValueStreamCode	.089	Ja

H2 One-Sample Test

H2.1 resultaten representatievoorkeur

Voor het vaststellen of de nieuwe versie intuïtiever is dan de oorspronkelijke versie is de *One-Sample Test* toegepast. In tabel 74 en tabel 75 zijn de persoonlijke resultaten weergegeven voor de representatievoorkeursvragen.

Tabel 74: Representatievoorkeursvragen: persoonlijke resultaten One-Sample Test (Statistics)

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
InternalGoalCode	14	53.21	26.718	7.141
ValueStreamCode	14	41.43	36.132	9.657

Tabel 75: Representatievoorkeursvragen: persoonlijke resultaten One-Sample Test

One-Sample Test						
Test Value = 0						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
InternalGoalCode	7.452	13	.000	53.214	37.79	68.64
ValueStreamCode	4.290	13	.001	41.429	20.57	62.29

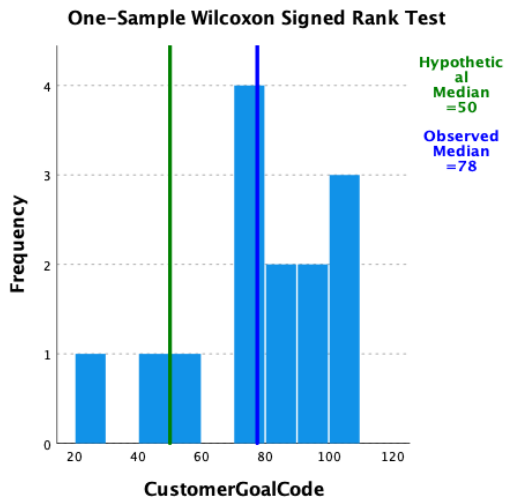
H3 One-Sample Wilcoxon Signed Rank Test

H3.1 Resultaat customer goal

Voor het vaststellen of de nieuwe versie van het PGA-element value proposition intuïtiever is dan de oorspronkelijke versie is de *One-Sample Wilcoxon Signed Rank Test toegepast*. In tabel 77 en figuur 4 zijn de persoonlijke resultaten weergegeven.

Tabel 76: Representatievoorkeursvragen: customer goal

One-Sample Wilcoxon Signed Rank Test Summary	
Total N	14
Test Statistic	86.000
Standard Error	14.199
Standardized Test Statistic	2.852
Asymptotic Sig.(2-sided test)	.004



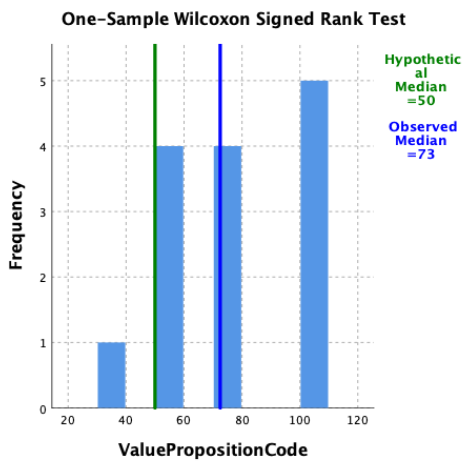
Figuur 4: Representatievoorkeursvragen: CustomergoalCode

H3.2 Resultaat value proposition

Voor het vaststellen of de nieuwe versie van het PGA-element value proposition intuïtiever is dan de oorspronkelijke versie is de *One-Sample Wilcoxon Signed Rank Test* toegepast. In tabel 77 en figuur 5 zijn de persoonlijke resultaten weergegeven.

Tabel 77: Representatievoorkeursvragen: value proposition

Total N	14
Test Statistic	64.000
Standard Error	11.124
Standardized Test Statistic	2.787
Asymptotic Sig.(2-sided test)	.005



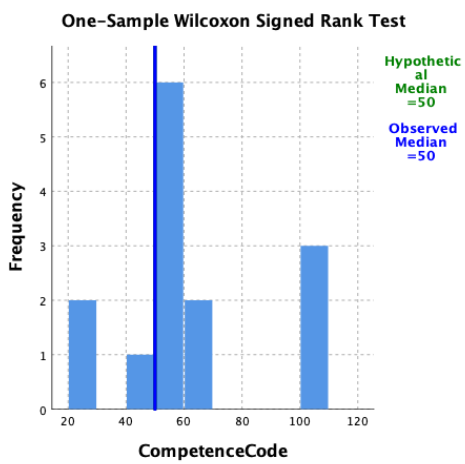
Figuur 5: Representatievoorkeursvragen: ValuePropositionCode

H3.3 Resultaat competence

Voor het vaststellen of de nieuwe versie van het PGA-element value proposition intuïtiever is dan de oorspronkelijke versie is de *One-Sample Wilconxon Signed Rank Test toegepast*. In tabel 78 en figuur 6 zijn de persoonlijke resultaten weergegeven.

Tabel 78: Representatievoorkeursvragen: competence

One-Sample Wilcoxon Signed Rank Test Summary	
Total N	14
Test Statistic	26.000
Standard Error	7.098
Standardized Test Statistic	1.127
Asymptotic Sig.(2-sided test)	.260



Figuur 6: Representatievoorkeursvragen: CompetenceCode

Bijlage L: Groepsresultaten van de representatievoorkeursvragen

H1 Normale verdeling

H1.1 resultaten van de vragen

Voor het vaststellen of een variabele normaal verdeeld is, is de *Shapiro-Wilk* toegepast. In tabel 79 de groepsresultaten weergegeven voor de representatievoorkeursvragen.

Tabel 79: Representatievoorkeursvragen: groepsresultaten Shapiro-Wilk test

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
CustomerGoalCode	.117	44	.153	.942	44	.028
ValuePropositionCode	.157	44	.008	.896	44	.001
sCompetenceCode	.179	44	.001	.930	44	.011
InternalGoalCode	.098	44	.200*	.964	44	.182
ValueStreamCode	.112	44	.200*	.921	44	.005

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

H1.2 Samenvatting van de representatievoorkeursvragen

In tabel 80 is de samenvatting weergegeven van de representatievoorkeursvragen.

Tabel 80: representatievoorkeursvragen resultaten: groepsresultaten Shapiro-Wilk test

Representatievoorkeursvragen	Sig.	Normaal verdeeld
CustomerGoalCode	.028	Nee
ValuePropositionCode	.001	Nee
CompetenceCode	.011	Nee
InternalGoalCode	.182	Ja
ValueStreamCode	.005	Nee

H2 One-Sample Test

H2.1 resultaten internal goal

Voor het vaststellen of de nieuwe versie intuïtiever is dan de oorspronkelijke versie is de *One-Sample Test* toegepast. In tabel 81 en tabel 82 zijn de groepsresultaten weergegeven voor de representatievoorkeursvragen.

Tabel 81: Representatievoorkeursvragen: groepstestresultaten One-Sample Test (Statistics)

	One-Sample Statistics			
	N	Mean	Std. Deviation	Std. Error Mean
InternalGoalCode	44	53.52	27.840	4.197

Tabel 82: Representatievoorkeursvragen: groepstestresultaten One-Sample Test

One-Sample Test						
Test Value = 50						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
InternalGoalCode	.839	43	.406	3.523	-4.94	11.99

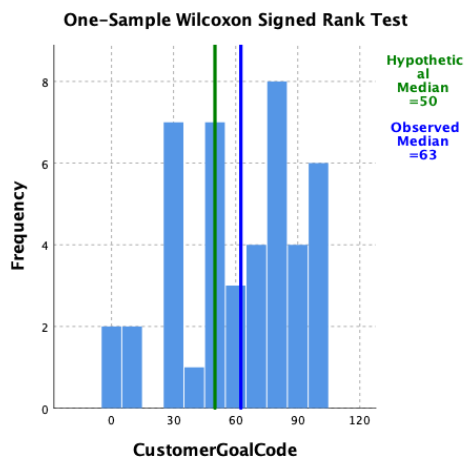
H3 One-Sample Wilcoxon Signed Rank Test

H3.1 Resultaat customer goal

Voor het vaststellen of de nieuwe versie van het PGA-element *customer goal* intuïtiever is dan de oorspronkelijke versie is de *One-Sample Wilcoxon Signed Rank Test* toegepast. In tabel 83 en figuur 7 zijn de persoonlijke resultaten weergegeven.

Tabel 83: Representatievoorkeursvragen: customer goal

One-Sample Wilcoxon Signed Rank Test Summary	
Total N	44
Test Statistic	486.000
Standard Error	65.999
Standardized Test Statistic	2.038
Asymptotic Sig.(2-sided test)	.042



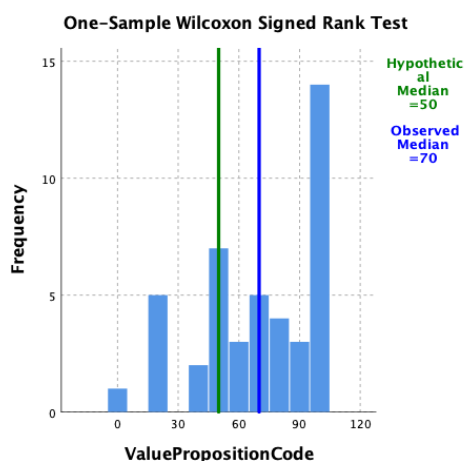
Figuur 7: Representatievoorkeursvragen: CustomerGoalCode

H3.2 Resultaat value proposition

Voor het vaststellen of de nieuwe versie van het PGA-element *value proposition* intuïtiever is dan de oorspronkelijke versie is de *One-Sample Wilcoxon Signed Rank Test* toegepast. In tabel 84 en figuur 8 zijn de groepsresultaten weergegeven.

Tabel 84: Representatievoorkeursvragen: value proposition

One-Sample Wilcoxon Signed Rank Test Summary	
Total N	44
Test Statistic	608.000
Standard Error	68.503
Standardized Test Statistic	3.467
Asymptotic Sig.(2-sided test)	.001



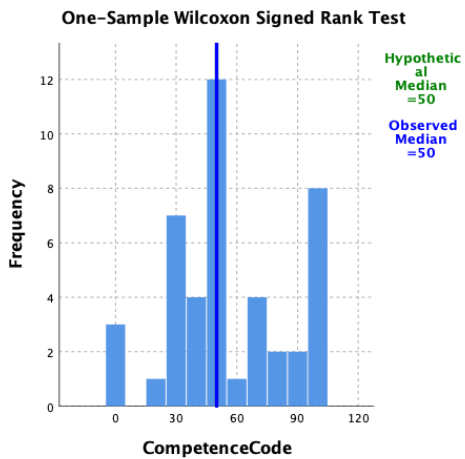
Figuur 8: Representatievoorkeursvragen ValuePropositionCode

H3.3 Resultaat competence

Voor het vaststellen of de nieuwe versie van het PGA-element *competence* intuïtiever is dan de oorspronkelijke versie is de *One-Sample Wilcoxon Signed Rank Test* toegepast in tabel 85 en figuur 9 zijn de groepsresultaten weergegeven.

Tabel 85: Representatievoorkeursvragen: Competence

One-Sample Wilcoxon Signed Rank Test Summary	
Total N	44
Test Statistic	342.000
Standard Error	55.645
Standardized Test Statistic	1.105
Asymptotic Sig.(2-sided test)	.269



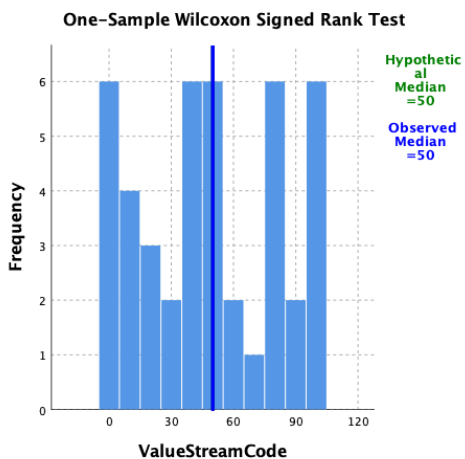
Figuur 9: Representatievoorkeursvragen: CompetenceCode

H3.4 Resultaat value stream

Voor het vaststellen of de nieuwe versie van het PGA-element *valuestream* intuïtiever is dan de oorspronkelijke versie is de *One-Sample Wilcoxon Signed Rank Test* toegepast. In tabel 86 en figuur 10 zijn de persoonlijke resultaten weergegeven.

Tabel 86: Representatievoorkeursvragen: value stream

One-Sample Wilcoxon Signed Rank Test Summary	
Total N	44
Test Statistic	335.000
Standard Error	68.649
Standardized Test Statistic	-.517
Asymptotic Sig.(2-sided test)	.605



Figuur 10: Representatievoorkeursvragen: ValueStreamCode

H4 resultaten toelichting op de representatievoorkeursvragen

In dit hoofdstuk zijn de toelichtingen opgenomen die bij de representatievoorkeursvragen zijn ingevuld. Als er 'no' de header van de tabel staat, dan betekent het dat links het nieuwe PGA-element is weergegeven en rechts het oorspronkelijke element. Voor 'on' is het andersom, dus links het oorspronkelijke element en rechts het nieuwe element. Het betreft de volgende vijf elementen:

1. Customer goal: zie tabel 87
2. Value proposition: zie tabel 88
3. Competente: zie tabel 89
4. Internal goal: zie tabel 90
5. Value stream: zie tabel 91

In de onderstaande tabellen is de toelichting voorzien van een score om zo een gedetailleerd beeld te krijgen van de feedback van de respondenten.

- Score 0: De toelichting geeft een algemene toelichting van de keuze
- Score 1: De toelichting geeft aan dat er een andere associatie is met de getoonde PGA-elementen
- Score 2: De toelichting geeft aan dat beide PGA-element niet intuïtief zijn

Tabel 87: Customer goal

CGNOClarification	Score	CGONClarification	Score
My association with Alternative B is more that of collaboration	1	Two hands look like an agreement between two people but could be two people in the company. When to persons shake hands and one has a suitcase it looks like a businessman agrees with a customer	1
Why does the customer carry a suitcase? How do I recognise the customer?	0	alternative A does not intuitively indicate a customer, can also be a business partner. alternative B icons indicate businessperson and non-businessperson	1
Bij A is een klant te zien. Bij B kan het ieder ander overeenkomst zijn.	1	Alternative B seems to be a more business agreement then A. The last one could also be another agreement.	1
alternative A might represent the concept of a visit, whereas alternative B is more clear about the concept of agreement or handshake, ie representing a desired outcome people are happy with	0	This picture makes the company / customer relationship more clear through the person with the suitcase.	0
Alt. A views a difference between the people due to the suitcase. so a customer-supplier relation is more obvious.	1	B introduces a briefcase that implies business, while A is cooperation in general.	1
easier to read	0	Both give me more of an impression that it is the sales goal, not the goal of the customer	1
aims more for working together	0	Alternative B shows 2 different stakeholders: a client and a salesman. That's a customer relationship.	0
A is more clear in terms of showing the relation between company and customer	0	Alternative B depicts a customer more clearly	0
Slight preference for A as it shows there is a customer (without briefcase) and a provider of services (with briefcase). However, both have a too strong association for me with a contractual agreement.	1	Just the two hands take too much space and make it a very busy icon. There is a little more whitespace in the two figures icon. I think the "goal" or "target" crosshairs use more screen real-estate in relation the real information in the icon. So both are bad imho.	2
none are very specific, looks more like collaboration	1	Showing persons instead of just the handshake, intuitively refers to real life customers	0
Seei v two persons, one with case	0	people shown: one with a case. suggesting business	1
No preference: in both images two people are coming to an agreement.	2	Easier to understand, both don't really represent a customer in my opinion	0
Alternative A looks more B2B	1	Alternative B indicates the agreement between people, whereas alternative A symbolizes cooperation in general.	1
Most simple	0	B does not look like a customer, but like a business partners	1
		I see at least one customer and one businessman	1
		Same for me, only zoomed in/out	2
		For me, alternative A stands for 'agreement' or 'deal', because of shaking hands.	1
		I find them both slightly off, more internal than external.	0

Tabel 88: Value proposition

VPNOClarification	Score	VPONClarification	Score
non are intuitive.	2	Alternative B: 2 people balancing a gift: they are balancing something. The scale in alternative A looks like there is balance: the action is missing	0
Alternative B is associated with justice	1	Alternative A indicates a balance, where B indicates a gift with a supposed value. Both are not ideal representations, yet B is a bit more intuitive	0
Ook hier klanten of anderen te zien. Weegschaal associeer ik direct met rechtspraak (beroepsdeformatie).	1	I would choose another symbol anyway (maybe something with a diamond). Alternative B looks a little bit similar as alternate for the customer goal	1
Alternative B seems to resemble more of dealing with legal disputes or a marketplace, whereas in A it is clear that something is offered/transferred from one entity to another	1	Both do not appeal.	2
universal understood icon	0	It gives some indication of the value of one thing to another, but it also does not fully cover it.	0
A does not make difference between the people icons. the product is too small to indicate it's about the product/services value. B is just no more than comparison. has no relation to product/services.	0	I don't associate the scale with value or proposition.	0
B reminds me of Justice	1	Even though value isn't necessary free of cost (which the present implies), it is at least related to delivering something. Undelivered value isn't value.	0
In A, the value is the present and the handing over of the present is the proposition. Weighing alternatives feels illogical: is there only 1 alternative compared to the value proposition? Typically, a customer must compare many value propositions.	0	The gift implies value for both parties (it's in the middle and could go either way). The scale implies a balance or law.	1
The gift is misleading	0	In picture 2 something is offered and in picture 1 something is being weighed on a scale	0
The 'value' is clearly represented by the gift in A.	0	Neither value is not best represented as a gift or a balance scale. I would prefer something like an upward trending graph.	2
Alternative A shows people caring, while Alternative B looks more like something that has to do with balance or law.	1	B shows actual value being 'presented'	0
The present is misleading	0	It's about the benefits. And that not just a product, it's a whole chain. The 'present picture' suggests a product to me	1
		The gift is not representing value, you want your customers to buy something, not gift something. The scales are not representing value to me	0
		Alternative B depicts the transfer of value to a customer, whereas alternative B symbolizes weighing something.	0
		A looks like balancing: what is best for the customer and for the P&L	0
		I see a proposing movement and something that is of value (the present). I do not see what a scale has to do with value and with a proposition	0
		Value is something you add or give to a person/product/service, people on the right icon give something	0
		Alternative B, because of the gift in the picture (I associate that more with value proposition)	0

Tabel 89: Competence

CNOClarification	Score	CONClarification	Score
In my opinion both alternatives don't represent competence. For me alternative A represents an idea and alternative B represents a competition not competence	2	Both icons are not very intuitive	2
Neither alternatives are directly associated with a competence.	2	Bij B denk ik meteen aan een persoon met ideeën, kennis en kunde. A associeer ik met het podium na een sportwedstrijd.	1
Both icons are not intuitive; alternative B indicates a contest and not a competence, alternative A has no link to competence for me	2	Alternative A makes me think of benchmarking or the Olympics; alternative B makes me think of ideas and using your brain	1
competence is not only about new ideas	0	A had no recognition with skills, just a visualization of rating B is somewhat closer to skills due to the icon of a person who has an idea.	0
Both do not represent competence. The first is about ideas, the second about winning.	2	Neither of the pics are really clear.	2
Most clear picto of the two. For me has alternative B a little match with the subject.	0	A feels better because you can excel in a competence. B is more associated with having a great idea. A is not optimal, however, as it indicates as if being the winner or scoring the result is the most important, whereas 'knowledge, skills, and abilities' seems to be the most important aspect of the definition of competence.	0
I associate A with ideas, not competence. B is associated with competition.	1	A only shows a ranking, while B shows it's about a person with certain knowledge/skills/competences.	0
A light bulb typically is related to an idea. Competence is something you are able to do. Competence is also not a competition, regardless of the similarity of the words. Both symbols are equally unsuitable. I would much rather see a diploma-like symbol or a person with a graduation hat.	1	Alternative A indicates being good at something although it redundantly incorporates a picture of competition and excellence. Alternative B strikes me as something that indicates innovation or a cognitive process, which both can be seen as a form of a competence but not as a general indicator of it.	1
Alternative A implies knowledge/idea while alternative b implies a comparison and the need of doing things well.	1	No difference	2
1 depicts having an idea and 2 shows being good at something	1		
The lamp bulb suggests invention. The first place suggest something you are good at. Still a bit far-fetched.	1		
Picture with 'lamp' suggest to me an idea. A 'superman' type (or winner) is more about what you can	1		
Both don't represent a set of knowledge, skills, and abilities to me, but B even less than A	1		
Alternative B symbolizes achievement in a specific area, whereas alternative A symbolizes a person's idea.	0		
Personal development versus being the best (in selling for example)	0		
Both not intuitive.	2		
Alternative B is about winning, alternative A is about a person with an idea. You could also use a picture of a person with an achievement ribbon to show his/her competence.	0		
Personally, my association with competences are related to people, so I tend to lean towards left	0		
Alternative A stands for idea.	1		

Tabel 90: Internal goal

IGNOClarification	Scope	IGONClarification	Score
In my opinion none of the alternatives represent internal goals. For me they both represent a process ore processes in the company	2	Both alternatives have the same association to me. A slight aesthetic preference for A	0
Both alternatives signal a goal for me, alternative A is more encompassing than B	0	Both icons are not very intuitive. The gear icon is usually used to indicate settings in a software program.	2
The first one represents more a business, the second a process, not a goal.	1	Bij B denk ik aan werkzaamheden in een fabriek.	1
The factory can seem confusing for me, but after explanation also useful.	0	The fact that the cog wheels are placed inside a factory clarifies that the goal is about something that happens inside an organization, ie better reflects the internal aspects of the goal	0
A is easier to understand as associated with something "within the factory".	0	B with the plant/factory icon within has less strategic value.	0
Alternative A at least puts the cog wheels inside a building, so it relates more towards internal goal. But I associate cog wheels with processes and tools, not with goals.	1	internal is better represented	0

IGNOClarification	Scope	IGONClarification	Score
The factory in alternative A is an internal process, while just the cogwheels in alternative B imply 'a proces' to me.	1	nicer	0
The factory is misleading for nonproduction goals	1	B looks like a factory	1
The factory does something indicate the cogwheels are internal. But again, it is way to busy, scale down on the crosshairs. I would prefer something like a simple arrow pointing into a circle.	1	Slight preference for A, because of clarity of the image. The icon could perhaps be more clear by showing the dot of the center of the target, and moving the cogs to one 'corner' of the image.	0
B skips the 'production' part that is shown in A (the factory). Personally, the factory is confusing to me....	0	B shows it's about the own factory/business, while A only shows movement.	1
The wheels to me indicate more of a process. The factory to me indicates internal workings	1	Very practical: the gears are better defined in Alternative A.	0
Not all companies are factories, but all companies have processes	0	The fabrique sign helps	0
Alternative A better depicts the process within a company, although it might be better visualized by a more general image of a company, e.g., an office building.	0		
don't like both of them, why a cogwheel in a goal symbol	2		
I associate the wheels with processing or producing. Alternative A is conveying too many messages to me, I find it confusing. I deduct that the circle should represent a goal or target, However, it is not what I see at first sight. The circular figure does not connect with any meaning or subject with me.	1		
Both don't really trigger my associations relating to goals	2		
I would prefer just the goal sign, without any images in the inner circle.	0		
Both not intuitive, seem to be about internal mechanics or operations management?	2		

Tabel 91: Value statement

VSONClarification	Score	VSNOClarification	Score
Alternative A for me is a very clear representation of this goal	0	The use of exclamation marks is very intuitive to express importance	0
The colors suggest something is wrong (red), okay (orange) and good (green).	0	Alternative A is more intuitive now that I know what it depicts.	0
Both alternatives are not intuitive for me as value streams; I associate a value stream with the porter model of layers with a point to the right.	2	De uitroeptekens vind ik verwarrend. Bij B zou ik andere of geen kleuren gebruiken (rood is meestal "verboden" of negatief/stop).	1
colours do better (otherwise too many !)	0	Alternative B looks more like different road types or indicating different types of optionality/degrees of freedom (because of the colors); it is clear the exclamation marks in alternative A represent importance	1
By far, the left one. The right one does not do anything for me. Colour is an important way of communicating importance.	0	A has most value for europeans who use 1 for less and 10 for highest scores. B is the other way around: red stands for low importance in fact not wanted, green for most wanted.	1
Easy to explain in a legend. Clear to use.	0	color coding is more intuitive	0
Colors are easier to read than counting exclamation marks.	0	more exclamation marks is more clear to me. red, yellow, green can mean important,, or good	1
Due to the colours, I had initially associated the red line with something bad and the green dotted line with something good. Usage of red and green is tricky given the percentage of people who are colour blind. As an alternative, I'd suggest ditching the colours from alternative A and adding the line weight of alternative B.	1	The colors implicate "bad (red)" or "good (green)". That is confusing.	1
red, yellow and green have connotations of a traffic light for me. Red = stop, Yellow = be alert, Green = everything is okay. The exclamation marks imply importance to me.	1	Preference for A, because the colors have a form of judgement: Red/orange is "wrong", whereas green means "good". The level of importance, however, should be neutral in my opinion.	1
Personally, I am more color oriented	0	The red color normally shows something is bad/wrong, while in this example it shows something positive (=of high importance). Therefore, B is counterintuitive.	1
Neither. Red is bad (although it indicates that it needs attention, but that does not imply importance in relation to the whole). The other one is, again, to busy.	2	Alternative B had more distinction between the types of importance due to use of color and line type.	0
Colors are better than symbols i.m.o. but the exclamation marks confer urgency and importance	0	The left one is more clear	0

VSONClarification	Score	VSNOClarification	Score
Dashes need explanation but keep a picture more clear. I think the exclamation marks are more intuitive. Yet I prefer dashes :)	0		
Colors are beter (for the ones that can see them properly). Form is also better than extra symbols	0		
Alternative A: the red line is generally associated with problems, and there is no intuitive distinction between the yellow and green dotted lines. Alternative B: the exclamation marks better indicate the relative importance of a relation line.	1		
The use of colors connects with safe/allowed (green), attention (yellow) and red (critical/forbidden). I prefer alternative B with the thickness of the lines, but without the exclamation marks as I interprete these as warnings! You might consider stars instead	1		
A: dots looks like a planning, colours indicate good or wrong	1		
Working with colors, strong association with urgency	0		
Alternative A, because of the colours.	0		

Bijlage M: Persoonlijke resultaten van de elementen geaggregeerd.

H1 Normale verdeling

H1.1 resultaten van de geaggregeerde elementen

Voor het vaststellen of een variabele normaal verdeeld is, is de *Shapiro-Wilk* toegepast. In tabel 92 zijn de persoonlijke resultaten weergegeven voor de representatievoorkeursvragen.

Tabel 92: Elementen geaggregeerd: persoonlijke resultaten Shapiro-Wilk test

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
GeaggregeerdInternalGoal	.169	14	.200*	.962	14	.754
GeaggregeerdValueProposition	.135	14	.200*	.954	14	.623
GeaggregeerdCompetence	.119	14	.200*	.959	14	.708
GeaggregeerdValueStream	.229	14	.045	.834	14	.013
GeaggregeerdCustomerGoal	.155	14	.200*	.966	14	.815

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

H1.2 Samenvatting van de representatievoorkeursvragen

In tabel 93 is de samenvatting weergegeven van de geaggregeerde elementen. Als sig. < is dan 0.05 dan is de nulhypothese weerlegd en is de conclusie dat de verdeling afwijkend is van de normale verdeling.

Tabel 93: Elementen geaggregeerd: groepsresultaten Shapiro-Wilk test

Representatievoorkeursvragen	Sig.	Normaal verdeeld
GeaggregeerdInternalGoal	.754	Ja
GeaggregeerdValueProposition	.623	Ja
GeaggregeerdCompetence	.708	Ja
GeaggregeerdValueStream	.013	Nee
GeaggregeerdCustomerGoal	.815	Ja

H2 Mann-Whitney U test

H2.1 Resultaten van de geaggregeerde elementen

Voor het vaststellen of de nieuwe versie beter is dan de oorspronkelijke versie is de *Mann-Whitney U Test* toegepast. In tabel 94 en tabel 95 zijn de persoonlijke resultaten weergegeven voor de geaggregeerde elementen.

Tabel 94: Elementen geaggregeerd: persoonlijke resultaten Mann-Whitney U test (Ranks)

	Ranks			
	VersionCode	N	Mean Rank	Sum of Ranks
GeaggregeerdValueStream	0	7	8.79	61.50
	1	7	6.21	43.50
	Total	14		

Tabel 95: Elementen geaggregeerd: persoonlijke resultaten Mann-Whitney U test (Test Statistics)

Test Statistics ^a		GeaggregeerdValueStream
Mann-Whitney U		15.500
Wilcoxon W		43.500
Z		-1.151
Asymp. Sig. (2-tailed)		.250
Exact Sig. [2*(1-tailed Sig.)]		.259 ^b

a. Grouping Variable: VersionCode

b. Not corrected for ties.

H3 Independent Samples Test

H3.1 Resultaten van de geaggregeerde elementen

Voor het vaststellen of de nieuwe versie beter is dan de oorspronkelijke versie is de *Independent Samples Test* toegepast. In tabel 96 en tabel 97 zijn de persoonlijke resultaten weergegeven voor de geaggregeerde elementen.

Tabel 96: Elementen geaggregeerd: persoonlijke resultaten Independent Samples Test (Group Statistics)

Group Statistics					
	VersionCode	N	Mean	Std. Deviation	Std. Error Mean
GeaggregeerdInternalGoal	0	7	261.4286%	46.71609%	17.65702%
	1	7	277.2222%	43.78093%	16.54763%
GeaggregeerdValueProposition	0	7	210.9524%	43.43852%	16.41822%
	1	7	199.3651%	104.58148%	39.52809%
GeaggregeerdCompetence	0	7	239.6825%	64.88186%	24.52304%
	1	7	191.5873%	47.33041%	17.88921%
GeaggregeerdCustomerGoal	0	7	259.4444%	24.98147%	9.44211%
	1	7	248.9683%	75.25774%	28.44475%

Tabel 97: Elementen geaggregeerd: persoonlijke resultaten Independent Samples Test (Independent Samples Test)

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
GeaggregeerdInternalGoal	Equal variances assumed	.068	.799	-.653	12	.526	-15.79365%	24.19906%	-68.51888%	36.93158%
	Equal variances not assumed			-.653	11.950	.526	-15.79365%	24.19906%	-68.54344%	36.95614%

GeaggregeerdValueProposition	Equal variances assumed	10.119	.008	.271	12	.791	11.58730%	42.80219%	-	104.84526%
									81.67066%	
	Equal variances not assumed			.271	8.010	.793	11.58730%	42.80219%	-	110.26701%
									87.09240%	
GeaggregeerdCompetence	Equal variances assumed	.551	.472	1.584	12	.139	48.09524%	30.35463%	-	114.23229%
									18.04181%	
	Equal variances not assumed			1.584	10.977	.141	48.09524%	30.35463%	-	114.92276%
									18.73228%	
GeaggregeerdCustomerGoal	Equal variances assumed	5.870	.032	.350	12	.733	10.47619%	29.97094%	-	75.77726%
									54.82488%	
	Equal variances not assumed			.350	7.306	.737	10.47619%	29.97094%	-	80.74817%
									59.79579%	

Bijlage N: Groepsresultaten van de elementen geaggregeerd

H1 Normale verdeling

H1.1 resultaten van de geaggregeerde elementen

Voor het vaststellen of een variabele normaal verdeeld is, is de *Shapiro-Wilk* toegepast. In tabel 98 zijn de groepsresultaten weergegeven voor de representatievoorkeursvragen.

Tabel 98: Elementen geaggregeerd: groepsresultaten Shapiro-Wilk test

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Geaggregeerd Internal Goal	.171	44	.002	.882	44	.000
Geaggregeerd Value proposition	.118	44	.141	.978	44	.559
Geaggregeerd Competence	.061	44	.200*	.984	44	.776
Geaggregeerd Value Stream	.170	44	.003	.936	44	.016
Geaggregeerd Customer Goal	.160	44	.007	.929	44	.010

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

H1.2 Samenvatting van de representatievoorkeursvragen

In tabel 99 is de samenvatting weergegeven van de geaggregeerde elementen. Als sig. < is dan 0.05 dan is de nulhypothese weerlegd en is de conclusie dat de verdeling afwijkend is van de normale verdeling.

Tabel 99: Elementen geaggregeerd: groepsresultaten Shapiro-Wilk test

Representatievoorkeursvragen	Sig.	Normaal verdeeld
GeaggregeerdInternalGoal	.000	Nee
GeaggregeerdValueProposition	.559	Ja
GeaggregeerdCompetence	.776	Ja
GeaggregeerdValueStream	.016	Nee
GeaggregeerdCustomerGoal	.010	Nee

H2 Mann-Whitney U test

H2.1 Resultaten van de geaggregeerde elementen

Voor het vaststellen of de nieuwe versie beter is dan de oorspronkelijke versie is de *Mann-Whitney U Test* toegepast. In tabel 100 en tabel 101 zijn de groepsresultaten weergegeven voor de geaggregeerde elementen.

Tabel 100 :Elementen geaggregeerd: groepsresultaten Mann-Whitney U test (Ranks)

	Ranks			
	VersionCode	N	Mean Rank	Sum of Ranks
Geaggregeerd Internal Goal	0	24	21.79	523.00
	1	20	23.35	467.00
	Total	44		
Geaggregeerd Value Stream	0	24	24.73	593.50
	1	20	19.83	396.50
	Total	44		
Geaggregeerd Customer Goal	0	24	23.08	554.00
	1	20	21.80	436.00
	Total	44		

Tabel 101: Elementen geaggregeerd: groepsresultaten Mann-Whitney U test (Test Statistics)

	Test Statistics ^a		
	Geaggregeerd Internal Goal	Geaggregeerd Value Stream	Geaggregeerd Customer Goal
Mann-Whitney U	223.000	186.500	226.000
Wilcoxon W	523.000	396.500	436.000
Z	-.401	-1.262	-.330
Asymp. Sig. (2-tailed)	.689	.207	.741

a. Grouping Variable: VersionCode

H3 Independent Samples Test

H3.1 Resultaten van de geaggregeerde elementen

Voor het vaststellen of de nieuwe versie beter is dan de oorspronkelijke versie is de *Independent Samples Test* toegepast. In tabel 102 en tabel 103 zijn de groepsresultaten weergegeven voor de geaggregeerde elementen.

Tabel 102: Elementen geaggregeerd: groepsresultaten Independent Samples Test (Group Statistics)

	Group Statistics				
	VersionCode	N	Mean	Std. Deviation	Std. Error Mean
Geaggregeerd Value proposition	0	24	201.7361%	61.92461%	12.64031%
	1	20	180.1667%	88.92837%	19.88499%
Geaggregeerd Competence	0	24	214.5602%	71.37067%	14.56848%
	1	20	195.8889%	71.87080%	16.07080%

Tabel 103: Elementen geaggregeerd: groepsresultaten Independent Samples Test (Independent Samples Test)

		Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Geaggregeerd Value proposition	Equal variances assumed	4.539	.039	.945	42	.350	21.56944%	22.81305%	-24.46915%	67.60804%
	Equal variances not assumed			.915	33.005	.367	21.56944%	23.56247%	-26.36847%	69.50736%
Geaggregeerd Competence	Equal variances assumed	.046	.832	.861	42	.394	18.67130%	21.67718%	-25.07502%	62.41761%
	Equal variances not assumed			.861	40.477	.394	18.67130%	21.69127%	-25.15229%	62.49489%

Bijlage O: Analyseprocedure

Deze bijlage beschrijft de aanpak die gevolgd is om de hypothesen te toetsen en welke stappen zijn gevolgd om te aggregeren over taken heen.

Test representatievoorkeursvragen

Als de *Mean rank* > 50 (waargenomen effect gaat in de richting van de alternatieve hypothese) dan:

- eenzijdige *p-value* = tweezijdige *p-value* / 2.
 - Als eenzijdige *p-value* ≤ 0.05,
 - dan H0 verwerpen en H1 aanvaarden (voorkeur voor de nieuwe notatie)
 - Als eenzijdige *p-value* > 0.05,
 - dan H0 niet verwerpen (geen duidelijk significant verschil in voorkeur tussen de oorspronkelijke en nieuwe notatie) en H1 niet aanvaarden.

Als *Mean rank* ≤ 50 (waargenomen effect gaat in tegenovergestelde richting van de alternatieve hypothese) dan:

- eenzijdige *p-value* = 1 – (tweezijdige *p-value*/2).
 - Als eenzijdige *p-value* ≤ 0.05,
 - dan H0 verwerpen en H1 aanvaarden (voorkeur voor de nieuwe notatie)
 - Als eenzijdige *p-value* > 0.05,
 - dan H0 niet verwerpen (geen duidelijk significant verschil in voorkeur tussen de oorspronkelijke en nieuwe notatie) en H1 niet aanvaarden.

Test overige vragen

Als *Mean rank* van de nieuwe versie > is dan *Mean rank* van de oorspronkelijke versie (waargenomen effect gaat in de richting van de alternatieve hypothese) dan:

- eenzijdige *p-value* = tweezijdige *p-value*/2.
 - Als eenzijdige *p-value* ≤ 0.05,
 - dan H1 aanvaarden (voorkeur voor de nieuwe notatie)
 - Als eenzijdige *p-value* > 0.05,
 - dan H0 niet verwerpen

Als *Mean rank* van de nieuwe versie ≤ is dan *Mean rank* van de oorspronkelijke versie (waargenomen effect gaat in tegenovergestelde richting van de alternatieve hypothese) dan:

- eenzijdige *p-value* = 1 – (tweezijdige *p-value*/2).
 - Als eenzijdige *p-value* ≤ 0.05
 - dan H1 aanvaarden (voorkeur voor de nieuwe notatie)
 - Als eenzijdige *p-value* > 0.05
 - dan H0 niet verwerpen

Test van de tijden

Als *Mean rank* van de oorspronkelijke versie > is dan *Mean rank* van de nieuwe versie (waargenomen effect gaat in de richting van de alternatieve hypothese) dan:

- eenzijdige *p-value* = tweezijdige *p-value*/2.
 - Als eenzijdige *p-value* ≤ 0.05
 - dan H1 aanvaarden (voorkeur voor de nieuwe notatie)
 - Als eenzijdige *p-value* > 0.05,
 - dan H0 niet verwerpen

Als *Mean rank* van de oorspronkelijke versie ≤ is dan *Mean rank* van de nieuwe versie (waargenomen effect gaat in tegenovergestelde richting van de alternatieve hypothese) dan:

- eenzijdige *p-value* = 1 – (tweezijdige *p-value*/2).
 - Als eenzijdige *p-value* ≤ 0.05
 - dan H1 aanvaarden (voorkeur voor de nieuwe notatie)

- Als eenzijdige p -value > 0.05
 - dan H_0 niet verwerpen

Aggregeren over taken heen

- Uitgangspunten
 - Analyse per PGA-element (want combinaties van oud & nieuw per respondent)
 - Behandel alle respondenten gelijk
 - Gewichten per taak kunnen, maar zijn niet verplicht (voorzie een goede argumentatie)
- Herschaal behaalde scores:
 - Voor *case study*: gemiddelde van vraag over *value stream* en *importance*
 - Voor notatie associatie betekent dit dat de behaalde scores met 1 verhoogd moeten worden: zo krijgen we enkel positieve scores
 - Representatievoorkeursvragen:
 - Indien een respondent in de voorgaande taken de nieuwe notatie van een PGA-element zag, kies de bestaande codering
 - Indien een respondent in de voorgaande taken de oude notatie van een PGA-element zag, nieuwe score = 100 - bestaande score
- Bepaal de maximale score voor een bepaalde taak
 - Notatie-associatievragen = 3
 - Begripsvragen = 1 (met uitzondering van value stream = 2)
 - Herinneringsvragen = 1
 - Representatievoorkeursvragen = 100
- Druk de behaalde scores uit als een percentage
- Tel de percentages op over taken heen (de resultaten zitten in een range tussen 0-400%)
- Hypotheses:
 - H_0 : de geaggregeerde score van de nieuwe notatie van PGA-element X is gelijk aan die van de oorspronkelijke notatie
 - H_1 : de geaggregeerde score van de nieuwe notatie van PGA-element X is groter dan die van de oorspronkelijke notatie
- Ga naar SPSS
 - Test normaliteit
 - Kies overeenkomstige test:
 - *T-test (unpaired)*
 - *Wilcoxon-Mann-Whitney-test*
 - Analyseer de resultaten en trek een conclusie m.b.t. hypothesen

Getoonde PGA-element per goep

Tabel 104 geeft een overzicht van de PGA-elementen die gebruikt zijn in de taken per groep: 0/A of 1/B. Dit is om te kunnen achterhalen welke versie van het PGA-element de participanten mee hebben gewerkt.

Tabel 104: Element volgorde in een taak

	Old	New
Version 0/A	Value proposition Value stream Customer goal Importance	Internal Goal Competence
Version 1/B	Internal Goal Competence	Value proposition Value stream Customer goal Importance