

De juiste wiskunde aanbevelen op digitale leeromgevingen: het effect van controle over aanbevolen oefeningen

Leen Dereu

Thesis voorgedragen tot het behalen
van de graad van Master of Science
in de ingenieurswetenschappen:
computerwetenschappen, hoofdoptie
Mens-machine communicatie

Promotor:

Prof. dr. Katrien Verbert

Co-promotor:

Dr. ir. Robin De Croon

Assessoren:

Katrien Laenen

Dr. ir. Jessa Bekker

Begeleider:

Jeroen Ooge

© Copyright KU Leuven

Zonder voorafgaande schriftelijke toestemming van zowel de promotor als de auteur is overnemen, kopiëren, gebruiken of realiseren van deze uitgave of gedeelten ervan verboden. Voor aanvragen tot of informatie i.v.m. het overnemen en/of gebruik en/of realisatie van gedeelten uit deze publicatie, wend u tot het Departement Computerwetenschappen, Celestijnenlaan 200A bus 2402, B-3001 Heverlee, +32-16-327700 of via e-mail info@cs.kuleuven.be.

Voorafgaande schriftelijke toestemming van de promotor is eveneens vereist voor het aanwenden van de in deze masterproef beschreven (originele) methoden, producten, schakelingen en programma's voor industrieel of commercieel nut en voor de inzending van deze publicatie ter deelname aan wetenschappelijke prijzen of wedstrijden.

Voorwoord

Deze thesis is het sluitstuk van vijf jaar studeren aan de KU Leuven. Ik heb een heel jaar lang kunnen werken aan een onderzoek dat me enorm boeide en op bepaalde momenten ook echt uitdaagde. Het finaliseren van de thesis was niet mogelijk geweest zonder de steun van een aantal mensen.

Allereerst wil ik Jeroen Ooge bedanken, zonder hem had het resultaat nooit hetzelfde geweest. Hij daagde me telkens uit om verder na te denken, dan ik soms in eerste instantie deed, over bepaalde keuzes of problemen. Daarnaast wil ik mijn promotor Prof. dr. Katrien Verbert en co-promotor Dr. ir. Robin De Croon bedanken voor de feedback na de tussentijdse presentaties. Vervolgens wil ik ook alle leerkrachten bedanken die deelnamen aan de pilootstudie en/of hun klas(sen) ter beschikking stelden voor het onderzoek. Tot slot is het ook belangrijk om mijn ouders en grootvader te bedanken om deze thesis kritisch na te lezen en/of de *think-aloud* studies te testen voordat ik ze uitvoerde bij het doelpubliek.

Leen Dereu

Inhoudsopgave

Voorwoord	i
Samenvatting	iv
Summary	v
1 Inleiding	1
2 Literatuurstudie	3
2.1 Pedagogische achtergrond	3
2.2 Digitale leeromgevingen	6
2.3 Aanbevelingssystemen	8
2.4 Feedback in digitale leeromgevingen	13
2.5 Gerelateerd werk	15
3 Methodes	17
3.1 Digitale leeromgeving Wiski	17
3.2 Pilootstudie	19
3.3 Deelnemers gebruikersstudie	20
3.4 Gerandomiseerd onderzoek met controlegroep	21
3.5 Oefeningen aanbevelen	23
3.6 Vragenlijst gebruikt na het onderzoek	24
3.7 Statistische analyse	25
4 Ontwikkeling	29
4.1 Pilootstudie	29
4.2 <i>Low-fidelity</i> prototype	31
4.3 <i>High-fidelity</i> prototype	41
4.4 Finaal <i>proof of concept</i>	46
4.5 Technische implementatie	48
5 Resultaten	51
5.1 Vertrouwen in de aanbevelingen	52
5.2 Interactie-adequaatheid	55
5.3 Waargenomen gebruiksgemak	56
5.4 Kwalitatieve data	58
6 Discussie	63
6.1 Waargenomen gebruiksgemak	63
6.2 Interactie-adequaatheid	63

6.3	Vertrouwen in de aanbevelingen	64
6.4	Digitale leeromgevingen in het onderwijs	67
6.5	Beperkingen van het onderzoek	68
7	Conclusie	71
7.1	Antwoorden op onderzoeksvragen	71
7.2	Mogelijkheden voor toekomstig onderzoek	72
A	Pilootstudie	75
A.1	Algemene vragen	75
A.2	Reeksen	75
A.3	Resultaten	76
B	Aanwervingsdocumenten gebruikersstudie	79
C	<i>Think-aloud</i> studies	85
C.1	Eerste <i>think-aloud</i> studie	85
C.2	Tweede <i>think-aloud</i> studie	86
D	Vragenlijst	89
D.1	Demografie	89
D.2	Vertrouwen in de aanbevelingen	89
D.3	Controle en transparantie	90
D.4	interactie-adequaatheid en waargenomen gebruiksgemak	91
D.5	Controle en uitleg over impact ervan	92
E	Initiële moeilijkheid van oefeningen	93
	Bibliografie	97

Samenvatting

Educatieve aanbevelingssystemen proberen de best mogelijke aanbevelingen te leveren, maar om het nog te verbeteren kan er *Interactive Machine Learning* aan gekoppeld worden, waarbij feedback gevraagd wordt aan de gebruiker. In de literatuur zijn er twee leemtes in verband met feedback en/of digitale leeromgevingen. Enerzijds worden ze vaak samen onderzocht, maar de focus ligt meestal op het geven van feedback aan leerlingen. Feedback geven als gebruiker wordt ook onderzocht, maar weinig in de context van digitale leeromgevingen. Anderzijds focussen onderzoeken met een educatieve context vaak op studenten uit het hoger onderwijs. Om deze leemtes verder te verkennen, is deze thesis geschreven.

Het doel is om te onderzoeken wat de effecten zijn als gebruikers gedeeltelijke controle krijgt over de aanbevolen oefeningen en hun eigen niveau. Er wordt gekozen voor de term ‘controle’ omdat de term ‘feedback’ meer omvat dan wat de leerlingen tijdens het onderzoek kunnen doen.

Een bestaande digitale leeromgeving, Wiski [41], is uitgebreid tot een aanbevelingssysteem waarbij de gebruikers hun eigen startniveau kiezen en na elke reeks oefeningen gedeeltelijke controle krijgen over de volgende reeks aanbevelingen. Daarnaast geeft Wiski een uitleg over de impact van de controle op het niveau van de gebruiker. Via een gerandomiseerd onderzoek met controlegroep wordt onderzocht wat de invloed is van de functionaliteiten op waargenomen gebruiksgemak, interactie-adequaatheid en vertrouwen van de gebruiker in de aanbevelingen.

Er werd een gebruikersstudie uitgevoerd met 71 leerlingen uit het secundair onderwijs die een doorstroom richting volgen. De resultaten tonen dat de toevoeging van de functionaliteiten geen effect heeft op waargenomen gebruiksgemak. Het is dus niet moeilijker om de website te gebruiken en de resultaten van interactie-adequaatheid en vertrouwen zijn niet beïnvloed door het waargenomen gebruiksgemak. De resultaten toonden ook aan dat de toevoeging van beide functionaliteiten interactie-adequaatheid verhoogt. Verder bleek dat de toevoeging van de uitleg over de impact van de controle ten opzichte van enkel controle de interactie-adequaatheid niet verhoogt. Daarnaast tonen de resultaten dat de toevoeging van de controle over de aanbevelingen geen invloed heeft op het vertrouwen van de gebruikers in de aanbevelingen. Verder zorgt de toevoeging van controle en een uitleg over de impact ervan voor een verhoging van de vertrouwensaspecten welwillendheid, eendimensionaal vertrouwen en geloofsovertuiging. Wanneer enkel controle vergeleken wordt met controle en een uitleg over de impact ervan, is er een verhoging van dezelfde vertrouwensaspecten en bovendien van bekwaamheid.

Summary

Educational recommendation systems try to provide the best possible recommendations, but to improve it even more, it can be coupled with Interactive Machine Learning, which asks for feedback from the user. There are two shortcomings in the literature related to feedback and/or digital learning environments. On the one hand, they are often investigated together, but the focus is usually on providing feedback to learners. Providing feedback as a user is also investigated, but little in the context of digital learning environments. On the other hand, studies with an educational context often focus on higher education students. To further explore these shortcomings, this thesis was written.

The aim is to investigate what the effects are when users are given partial control over the recommended exercises and their own level. The term ‘control’ is chosen because the term ‘feedback’ encompasses more than what the students can do during the study.

An existing digital learning environment, Wiski [41], is extended to a recommendation system in which users choose their own starting level and after each set of exercises they have partial control over the next set of recommendations. In addition, Wiski explains the impact of the control on the user’s level. A randomized study with control group is used to investigate the impact of the functionalities on perceived ease of use, interaction adequacy and users’ trust in the recommendations.

A user study was conducted with 71 secondary school students following a theoretical orientation. The results show that the addition of the functionalities has no effect on perceived ease of use. Thus, it is not more difficult to use the website and the results of interaction adequacy and trust are not affected by perceived ease of use. The results also showed that the addition of both functionalities increases interaction adequacy. Furthermore, the results showed that the addition of the explanation of the impact of the control compared to just control does not increase interaction adequacy. In addition, the results show that the addition of control over recommendations does not affect users’ trust in the recommendations. Furthermore, the addition of control and an explanation of its impact increases the trust aspects of benevolence, one-dimensional trust, and trusting beliefs. When only control is compared to control and an explanation of its impact, there is an increase in the same trust aspects and, moreover, in competence.

Hoofdstuk 1

Inleiding

Aanbevelingssystemen zijn niet meer weg te denken uit het dagelijks leven, maar er zijn een aantal leemtes in de praktijk en in de literatuur. Tijdens dit onderzoek worden twee leemtes verder verkend in de context van digitale leeromgevingen. Ten eerste is er veel literatuur te vinden over feedback en digitale leeromgevingen, maar de focus ligt grotendeels op het geven van feedback aan de leerling [11, 12, 28, 54, 60]. Feedback geven als gebruiker komt ook voor, maar weinig in de context van digitale leeromgevingen [35, 59]. Ten tweede zijn er weinig studies die focussen op het secundair onderwijs [4, 41]: de meeste focussen op studenten uit het hoger onderwijs [1, 2, 9, 10, 11, 24, 28, 29, 33, 35, 45, 49, 54, 60].

Het doel is om te onderzoeken wat de effecten zijn als gebruikers gedeeltelijke controle krijgt over de aanbevolen oefeningen en hun eigen niveau. In deze thesis wordt de term ‘controle’ gebruikt omdat de term ‘feedback’ te breed is en dus meer omvat dan de controle die de leerlingen krijgen. De digitale leeromgeving Wiski [41] is uitgebreid tot een aanbevelingssysteem waarbij gebruikers hun eigen startniveau kiezen en na elke reeks oefeningen gedeeltelijke controle krijgen over de volgende reeks aanbevelingen. Daarbovenop geeft Wiski een uitleg over de impact van de controle op het niveau van de gebruiker. Er is gekozen voor een gerandomiseerd onderzoek met controlegroep met drie onderzoeksgroepen: (1) geen controle, (2) controle en (3) controle en een uitleg over de impact ervan. Met het onderzoek wordt de invloed onderzocht van de functionaliteiten op waargenomen gebruiksgemak, interactie-adequaatheid en het vertrouwen van de gebruiker in de aanbevelingen. Het onderzoek is relevant voor de domeinen mens-machine interactie en educatie, omdat het *Interactive Machine Learning* koppelt aan een aanbevelingssysteem, gebaseerd is op literatuur, een aantal leemtes verkent en digitale leeromgevingen steeds relevanter worden in de praktijk.

Hoofdstuk 2 bespreekt eerst de bestaande literatuur, zowel achtergrond informatie op vlak van educatie en technologie als gerelateerd werk. Daarna wordt in Hoofdstuk 3 de methode besproken die gevolgd is om het onderzoek tot stand te brengen. Hoofdstuk 4 bespreekt het ontwikkelingsproces van de digitale leeromgeving Wiski. Vervolgens toont Hoofdstuk 5 de resultaten van het onderzoek en de discussie wordt gevoerd in Hoofdstuk 6. Tot slot volgt de conclusie in Hoofdstuk 7.

Hoofdstuk 2

Literatuurstudie

Digitale leeromgevingen zijn de afgelopen jaren steeds relevanter geworden en de Covid-19 pandemie heeft het verder versterkt. Tijdens volledig of gedeeltelijk afstandsonderwijs mogen de pedagogische doelen niet over het hoofd gezien worden en aanbevelingssystemen kunnen hier een belangrijke ondersteunende rol spelen [21, 29]. Hierna wordt eerst de pedagogische achtergrond besproken, waarbij digitaal leren een belangrijke rol speelt. Daarna worden digitale leeromgevingen, aanbevelingssystemen, feedback binnen digitale leeromgevingen en gerelateerd werk behandeld.

2.1 Pedagogische achtergrond

Tijdens het ontwikkelen van een digitale leeromgeving is het belangrijk om de pedagogische doelen niet uit het oog te verliezen. Daarom bespreekt deze paragraaf waarop de eindtermen van het secundair onderwijs in België gebaseerd zijn, de rol die technologie kan spelen in het behalen van de eindtermen en het Dreyfus model voor het verwerven van vaardigheden [15].

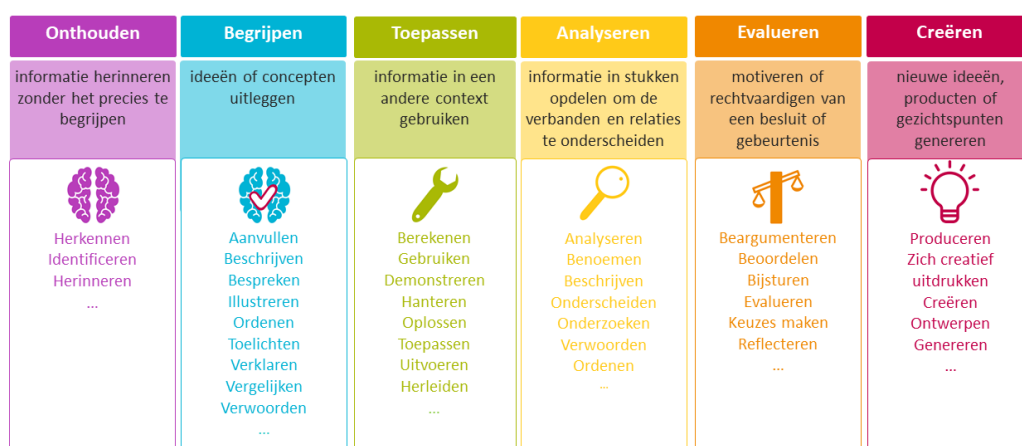
2.1.1 Herwerkte taxonomie van Bloom

De eindtermen van zowel het Gemeenschapsonderwijs als het Katholiek Onderwijs maken gebruik van de herwerkte taxonomie van Bloom, dat een kader biedt om leerdoelen te structureren [5, 25, 39]. Het maakt daarvoor twee belangrijke indelingen. De eerste indeling heeft te maken met vier kennisniveaus, feitenkennis, conceptuele kennis, procedurele kennis en metacognitieve kennis, die zijn samengevat in Tabel 2.1 [40]. De tweede indeling is op basis van de cognitieve dimensie, de affectieve dimensie en/of de psychomotorische dimensie. De cognitieve dimensie bestaat uit zes beheersingsniveaus die zijn samengevat in Figuur 2.1.

2. LITERATUURSTUDIE

TABEL 2.1: Soorten kennis volgens Onderwijs Vlaanderen [40].

Kennis	Beschrijving
Feitenkennis	“Deze kennis omvat de termen, begrippen en elementen die de leerlingen actief kunnen gebruiken om over een bepaald domein van gedachten te wisselen of om problemen binnen dat domein op te lossen.”
Conceptuele kennis	“Deze kennis omvat begrip en inzicht in classificaties, principes, theorieën en modellen die de leerlingen gebruiken bij het verwerken van andere kennis.”
Procedurele kennis	“Deze kennis omvat technieken, methoden en algoritmes ter ondersteuning van hoe de leerlingen iets uitvoeren, alsook van de criteria voor het kiezen van de geschikte procedure.”
Metacognitieve kennis	“Deze kennis omvat zelfkennis, kennis over kennis en strategische kennis die de leerlingen gebruiken om te reflecteren over zichzelf en het eigen leerproces.”



FIGUUR 2.1: Beheersingsniveaus cognitieve dimensie [38].

2.1.2 Digitale technologie en leren

Om de eindtermen te behalen, wordt geprobeerd digitale technologie meer in te schakelen. Volgens Lai [29] heeft technologie op dit moment vaak een ondersteunende rol en heeft het de fundamentele manier van leren niet veranderd. Een voorbeeld is het gebruik van een beamer en een digitale presentatie, het ondersteunt de leerkracht tijdens het lesgeven, maar verandert de manier van leren van een leerling niet. Het niet fundamenteel veranderen van de manier van leren, ligt volgens Lai [29] deels aan de weinige voordelen voor leerkrachten. Ze moeten veel inspanningen leveren om hun manier van lesgeven aan te passen, maar krijgen hiervoor persoonlijk niets in de plaats. Nochtans kan digitale technologie een belangrijke rol spelen in het behalen van de competentie om levenslang leren mogelijk te maken, wat in deze kennismaatschappij essentieel is. De vaardigheid hebben om ook in een informele context te kunnen leren moet ontwikkeld worden op de schoolbanken. Digitale leeromgevingen kunnen makkelijk op de schoolbanken en thuis ingezet worden, waardoor leerlingen het leren

in een informele context stapsgewijs onder de knie krijgen via een hybride vorm. Het voorgaande toont dat digitale leeromgevingen ook een pedagogisch nut kunnen hebben en naast technische kwaliteit moeten ontwikkelaars hier ook voor zorgen.

Om de competentie levenslang leren te behalen is het belangrijk om zelfregulerend leren onder de knie te hebben. Volgens van Der Veen et al. [53] betekent het: “Een geheel van kennis, vaardigheden en houdingen die leerlingen moeten verwerven om zelfstandig met vakken te kunnen werken en die kan worden overgezet van de ene leercontext naar de andere.” Zelfregulerend leren wordt ook uitgebreid besproken door B. J. Zimmerman [60] en volgens hem omvatten definities voor zelfregulerend leren drie kenmerken: zelfregulerende processen (metacognitief, motivationeel en gedragsproces, zie Tabel 2.2), self-oriented feedback loop waarbij leerlingen de effectiviteit van hun leermethode monitoren en hier ook uit leren, en een indicatie over hoe en waarom leerlingen een bepaalde strategie kiezen.

TABEL 2.2: Zelfregulerende leerprocessen [60].

Leerproces	Betekenis
Metacognitief proces	“Zelfregulerende leerlingen plannen, stellen doelen, organiseren, houden toezicht over zichzelf en evalueren zichzelf op verschillende momenten tijdens het proces”
Motivationeel proces	“De leerlingen melden een hoge mate van doeltreffendheid, zelf-toekenning, en intrinsieke taakinteresse”
Gedragsproces	“Zelfregulerende leerlingen selecteren, structureren en creëren omgevingen die het leren optimaliseren”

2.1.3 Dreyfus model

Het Dreyfus model is een model met vijf niveaus voor het verwerven van vaardigheden, oorspronkelijk ontwikkeld voor schaakspelers [15]. Een samenvatting van het model staat in Tabel 2.3, waarbij de namen van de niveaus de vertalingen zijn van Toolshero [57]. De tabel bestaat uit vijf kolommen die elk een deel van het niveau verduidelijken. (1) *Vaardigheidsniveau* omvat de verschillende niveaus van het model. (2) *Componenten* verwijst naar de situatie die een leerling waarneemt, dat kan contextvrij zijn en enkel betrekking hebben op algemene aspecten of situationeel en betrekking hebben op specifieke situaties. (3) *Perspectief* omvat het moment waarop leerlingen meer componenten beginnen te herkennen en ze moeten kiezen op welke de focus moet liggen. De leerlingen nemen zo een perspectief aan. (4) *Beslissing* omvat de manier waarop leerlingen beslissen om te handelen in een specifieke situatie, wat gebaseerd is op analytisch redeneren of een intuïtieve beslissing is. (5) *Betrokkenheid* beschrijft de mate waarin een leerling wordt ondergedompeld in de situatie.

TABEL 2.3: Vijf fasen model voor vaardigheidsverwerving door Dreyfus [15].

Vaardigheids-niveau	Componenten	Perspectief	Beslissing	Betrokkenheid
Beginner	Contextvrij	Geen	Analytisch	Terughoudend
Gevorderd	Contextvrij, situationeel	Geen	Analytisch	Terughoudend
Competent	Contextvrij, situationeel	Gekozen	Analytisch	Terughoudend begrijpen en beslissen; Betrokken uitkomst
Bedreven	Contextvrij, situationeel	Ervaren	Analytisch	Betrokken begrijpen; Terughoudend beslissen
Expert	Contextvrij, situationeel	Ervaren	Intuïtief	Betrokken

De vaardigheidsniveaus kunnen gebruikt worden in een educatieve context en dan houden ze het volgende in:

- **Beginner.** De leerling kent de regels en feiten verbonden aan het onderwerp, maar heeft weinig tot geen kennis over de specifieke context.
- **Gevorderd.** Na voldoende voorbeelden herkent de leerling patronen.
- **Competent.** De leerling herkent nu zoveel patronen, dat hij of zij het overzicht kwijt is. Om competent te worden moet de leerling leren om zich te beperken tot de relevante patronen.
- **Bedreven.** De leerling ziet onmiddellijk het doel, maar weet nog niet onmiddellijk hoe het behaald kan worden.
- **Expert.** De leerling ziet nu ook onmiddellijk hoe het doel behaald kan worden.

2.2 Digitale leeromgevingen

Er bestaan verschillende methodes waarmee digitale leeromgevingen het niveau van een leerling kunnen bepalen en aanpassen. In deze paragraaf worden twee voorbeelden besproken, namelijk *Item Response Theory* en het Elo-classificatiesysteem.

2.2.1 *Item Response Theory*

Item Response Theory (IRT) kan gebruikt worden om het vaardigheidsniveau van een persoon te bepalen aan de hand van de prestaties van die persoon [24, 27, 46]. Het werd oorspronkelijk ontwikkeld voor psychometrische studies en het stelt dat de waarschijnlijkheid van een specifieke respons wordt bepaald door de eigenschappen van het item en de aard van de studiepersoon [19]. IRT wordt meestal toegepast in het *Computerized Adaptive Testing* (CAT) domein om de best passende items voor leerlingen te selecteren gebaseerd op individuele vaardigheden [9, 24, 27, 59]. Maths Garden gebruikt het bijvoorbeeld voor het monitoren van wiskunde in het basisonderwijs [27]. Bijna alle IRT-modellen zijn variaties op het simpelste model, het Rasch-model [24, 27, 59], waar de kans op een juist antwoord een logistische functie is van het verschil tussen de vaardigheid van de leerling en de moeilijkheid

van het item. Het verschil wordt gewogen met een parameter die een hoge waarde heeft wanneer een item een goed onderscheid maakt tussen laag- en hoogbegaafde leerlingen [27, 46, 59].

Er zijn echter een aantal problemen bij het gebruik van IRT voor CAT-platformen [27, 46, 59]. Ten eerste moet de moeilijkheidsgraad van de items vooraf gekend zijn, hiervoor moet een grote representatieve steekproef van de doelgroep de items beantwoorden, wat tijdrovend en kostbaar is [27]. Ten tweede werkt CAT het meest efficiënt als de moeilijkheidsgraad van de aanbevolen items gelijk is aan de inschatting van het vaardigheidsniveau van de persoon [27, 59]. Het meest informatieve item heeft dan een slagingspercentage van 50%, wat als demotiverend aanvoelt voor leerlingen. Onderzoek door Eggen et al. [17] toonde aan dat een verhoging van het percentage tot boven 70% de nauwkeurigheid sterk verlaagd. Stel dat er gestreefd wordt naar een percentage van 70%, dan moeten er voor een nauwkeurige schatting dus meer items aangeboden worden, wat de efficiëntie van CAT verlaagt. Ten derde moet er voor psychologische en educatieve metingen een afweging gemaakt worden tussen snelheid en nauwkeurigheid. Ten slotte gaan IRT-modellen er meestal vanuit dat de vaardigheid van een leerling constant blijft, omdat ze typisch gebruikt worden bij testen en er dan geen verandering is van niveau [46].

2.2.2 Elo-classificatiesysteem

De problemen van CAT en IRT worden vaak aangepakt door het Elo-classificatiesysteem te gebruiken. Het Elo-classificatiesysteem is oorspronkelijk ontwikkeld voor het classificeren van schakers en is gebaseerd op een gepaarde vergelijking van gegevens waarbij twee schakers het tegen elkaar opnemen [1, 18]. Het basisprincipe is eenvoudig: elke speler krijgt een niveau dat wordt herzien na elke match, proportioneel met hoe verrassend het resultaat van de match is [44, 46]. Als een sterke speler een zwakkere speler verslaat, is het resultaat niet verrassend en zal de verandering klein zijn; als de zwakkere speler wint, zal de verandering groot zijn.

Het Elo-classificatiesysteem wordt nu in meer detail besproken en is gelijkaardig aan het Rasch-model [44, 46]. Voor elke speler i is er een niveau schatting θ_i , $R_{ij} \in \{0, 1\}$ is het resultaat van de match tussen spelers i en j . De verwachte kans dat speler i wint is gegeven door een logistische functie met betrekking tot het verschil tussen de geschatte niveaus. De constante K , in de vergelijking die de niveaus van de spelers herziert (cfr. 2.2), geeft de gevoeligheid van de schatting van het niveau voor de laatste poging aan.

$$P(R_{ij} = 1) = \frac{1}{1 + e^{-(\theta_i - \theta_j)}} \quad (2.1)$$

$$\theta_i = \theta_i + K \times (R_{ij} - P(R_{ij} = 1)) \quad (2.2)$$

Om het Elo-classificatiesysteem te gebruiken in een educatief aanbevelingssysteem dienen er een aantal aanpassingen te gebeuren [46]. In een educatieve context wordt de leerling gezien als de speler en het item is de tegenstander. Het antwoord van een leerling op een item is dus een match tussen de leerling en het item [1, 27, 44, 46, 59].

Het niveau van een leerling s wordt weergegeven door θ_s en de moeilijkheid van het item i is d_i . De correctheid van het antwoord van leerling s op item i wordt weergegeven door $correct_{si} \in \{0, 1\}$. De kans op een correct antwoord wordt gegeven door een logistische functie met betrekking tot het verschil tussen het niveau van de leerling en de moeilijkheid van het item.

$$P(correct_{si} = 1) = \frac{1}{1 + e^{-(\theta_s - d_i)}} \quad (2.3)$$

Het niveau en de moeilijkheid worden daarna aangepast als volgt:

$$\theta_s = \theta_s + K \times (correct_{si} - P(correct_{si} = 1)) \quad (2.4)$$

$$d_i = d_i - K \times (correct_{si} - P(correct_{si} = 1)) \quad (2.5)$$

De parameter K weegt de impact van de afwijking van de verwachting tegenover het nieuwe niveau. Elo kan verandering van niveau volgen, waardoor het de voorkeur krijgt op IRT-modellen, die een vast niveau veronderstellen [46, 59]. In aanvulling is het ook mogelijk om 2.3 aan te passen zodat er rekening wordt gehouden met meerkeuzevragen. Een leerling heeft namelijk een significante kans om correct te antwoorden door te gokken. Voor een vraag met k opties, wordt de vergelijking aangepast als volgt:

$$P(correct_{si} = 1) = \frac{1}{k} + \frac{1 - \frac{1}{k}}{1 + e^{-(\theta_s - d_i)}} \quad (2.6)$$

Tot slot is het ook mogelijk om de tijd nodig om een vraag te beantwoorden in rekening te brengen [46].

In het klassieke Elo-classificatiesysteem wordt één parameter gebruikt om het niveau van de leerling te bepalen over het hele domein [1]. Voor een domein dat bestaat uit meerdere concepten bestaat er ook een variant, multi-concept multivariaat op Elo gebaseerd leermodel (M-Elo) [1], waarbij meerdere onafhankelijke parameters gebruikt worden om het niveau van de leerling te bepalen voor elk individueel concept.

2.3 Aanbevelingssystemen

Aanbevelingssystemen worden gedefinieerd als hulpmiddelen en technieken ontwikkeld om suggesties te doen over items die nuttig kunnen zijn voor een gebruiker [2, 8, 34, 35, 37, 47, 48]. De meeste aanbevelingssystemen kunnen onderverdeeld worden in vier domeinen: entertainment (bijvoorbeeld aanbevelingen voor films), inhoud (bijvoorbeeld toepassingen voor digitaal leren), e-commerce (bijvoorbeeld aanbevelingen om een boek te kopen) en services (bijvoorbeeld aanbevelingen voor reisbureaus) [48]. Het voordeel van aanbevelingssystemen is dat individuen die te weinig ervaring of kennis hebben om te kiezen uit een potentieel overweldigend aantal alternatieven begeleid worden bij hun keuze. Er zijn verschillende redenen waarom aanbevelingssystemen gebruikt worden door aanbieders van diensten: verhogen van aantal verkochte items, meer diverse items verkopen, verhogen van tevredenheid

gebruiker, verhogen van loyaliteit van gebruikers en beter begrijpen wat de gebruiker wil [48]. De redenen kunnen ook vertaald worden naar niet-commerciële toepassingen: verhogen van het aantal verkochte items kan bijvoorbeeld vertaald worden naar verhogen van het aantal gelezen artikels op een nieuwssite. Daarnaast zijn populaire taken waarbij aanbevelingssystemen kunnen assisteren te lezen in Tabel 2.4.

TABEL 2.4: Taken waarbij een aanbevelingssysteem kan assisteren [48].

Taken aanbevelingssysteem
Vinden van goede items.
Vinden van alle goede items.
Annotatie in context.
Aanbevelen van een reeks.
Aanbevelen van een bundel.
Gebruiker helpen om rond te kijken (zonder intentie om te kopen).
Gebruikersprofiel verbeteren.
Vinden van een geloofwaardige aanbeveler.
Gebruiker helpen zichzelf uit te drukken.
Gebruikers de mogelijkheid geven om anderen te helpen.
Gebruikers de mogelijkheid geven om anderen te beïnvloeden.

Aanbevelingstechnieken worden vaak onderverdeeld in vijf categorieën [8, 34, 48]:

- **Content-gebaseerd.** Er worden items aanbevolen die gelijkaardig zijn aan diegene die de gebruiker in het verleden leuk vond. Hiervoor worden twee bronnen gebruikt: de kenmerken van de items en de waardering van de gebruiker voor dat item.
- **Collaboratieve filtering.** De aanbevelingen worden gebaseerd op wat gelijkaardige gebruikers leuk vinden. Ze worden geclusterd door de gelijkenis van hun waarderingsgeschiedenis.
- **Demografisch.** Het demografisch profiel van de gebruiker wordt gebruikt om aanbevelingen te produceren.
- **Kennis-gebaseerd.** De items die aanbevolen worden voldoen aan de behoeften en voorkeuren van de gebruiker.
- **Hybride.** Een combinatie van de bovenstaande technieken.

Tot slot brengen aanbevelingssystemen een aantal uitdagingen met zich mee, de volgende opsomming is onvolledig [8, 48]. Ten eerste het *cold start* probleem voor nieuwe items en gebruikers, bijvoorbeeld bij collaboratieve filtering zijn de aanbevelingen van items met een lange waarderingsgeschiedenis beter onderbouwd dan aanbevelingen van nieuwe items. Ten tweede kan een systeem dat goed werkt met een kleine hoeveelheid informatie op grote schaal voor nieuwe problemen zorgen. Ten derde is het makkelijker voor een gebruiker om een gepast item te vinden als er een zekere diversiteit onder de aanbevolen items is. Ten vierde zou een systeem transparant moeten zijn door aan de gebruiker uit te leggen hoe het werkt. Ten vijfde moet er vertrouwen zijn in het systeem en in de andere gebruikers.

2.3.1 Educatieve aanbevelingssystemen

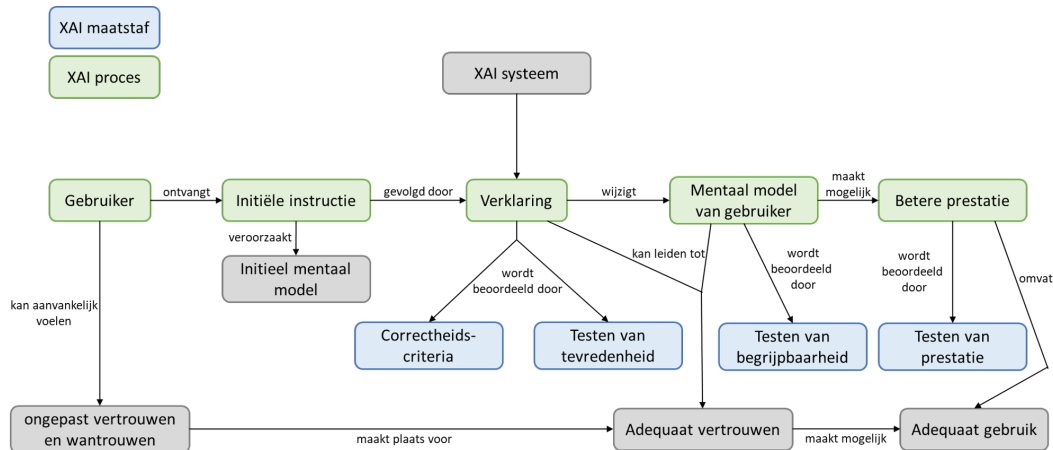
Educatieve aanbevelingssystemen zijn in bepaalde opzichten anders dan klassiek aanbevelingssystemen [21, 34]. Het doel is om de gebruiker te helpen bij het vinden van hulpmiddelen om het leerdoel makkelijker en sneller te behalen. Bovendien verschilt de context, omdat bij educatieve aanbevelingssysteem de context pedagogisch is, terwijl dat bij andere type aanbevelingssystemen vaak commercieel is. Verder zorgen pedagogische factoren ervoor dat het gebruikt wordt als hulpmiddel om het leerdoel makkelijker te bereiken en het is dus niet gebaseerd op persoonlijke voorkeuren. Daarnaast gebeurt de classificatie van de gebruiker op basis van de functie of het kennisniveau van de gebruiker. De leercyclus binnen educatieve aanbevelingssystemen omvat verschillende stappen: ontwerp van de leerervaring, beheer van de omgeving, gebruikstijd en feedback, leerproces en evaluatie. Ze gebruiken eveneens verschillende technieken, niet alleen collaboratieve filtering toepassen of verbeteren door rekening te houden met de bijzonderheden van leren maar ook alternatieven voor collaboratieve filtering worden gebruikt [14]. Als bron voor informatie worden vaak onderwijsbeperkingen gebruikt. Bovendien is het mogelijk om rekening te houden met de context van leren om het aanbevelingssysteem te verbeteren. Verder hebben ze verschillende ambities, zoals het onderwijseffect van de aanbevelingen beoordelen of lessen aanbevelen (i.p.v. de elementen erin).

De systemen met voorgaande kenmerken hebben verschillende doelen, namelijk extra leermiddelen voorstellen, het leerproces begeleiden of extra cursussen voorstellen. Daarnaast kunnen deze systemen ook de leerkracht ondersteunen om hun cursussen te verbeteren of hun leermiddelen te monitoren.

2.3.2 *Explainable Artificial Intelligence* (XAI)

Aanbevelingssystemen en hun uitdagingen passen in de bredere context van XAI omdat ze vaak een *black-box* zijn, er wordt een antwoord gegeven, maar er wordt niet uitgelegd waarom dat specifieke antwoord gegeven wordt [3, 42, 52]. De *black-box*-aanbevelingen kunnen leiden tot vertrouwensproblemen, vooral wanneer de gebruiker niet akkoord is met de aanbevelingen [2]. De Europese Unie [23] stelt het volgende: “Gebruikers hebben recht op een verklaring betreffende beslissingen genomen door een algoritme die gebaseerd zijn op persoonlijke informatie”. Volgens Hind [22] wordt er ook steeds vaker een verklaring bij het antwoord verwacht door de maatschappij, omdat ze een steeds grotere verspreiding hebben in de maatschappij [52]. De behoefte aan verklaringen kan voortkomen uit (ten minste) vier redenen: verklaren om te rechtvaardigen, verklaren om te controleren, verklaren om te verbeteren en verklaren om te ontdekken [3].

Het doel van *Explainable Artificial Intelligence* (XAI) is om verklaringen te bieden over de besluiten en acties van het systeem met het oog op meer transparantie en vertrouwen [3, 6]. Volgens Ooge et al. [42] kunnen de technieken voor verklaringen opgedeeld worden in drie categorieën, op basis van hoe ze het *black-box*-probleem openen: het model zelf uitleggen, de uitkomsten uitleggen of het model inspecteren. Een conceptueel model van de XAI-verklaringsprocedure staat in Figuur 2.2 [23].



FIGUUR 2.2: Een conceptueel model van de XAI-verklaringsprocedure [23].

Bovendien zijn er verschillende soorten verklaringen nodig voor vier soorten groepen gebruikers [22]. Ten eerste willen de ontwikkelaars van het systeem weten of het wel werkt zoals verwacht. Zo niet, willen ze het kunnen verbeteren met achtergrondkennis. Ten tweede de beleidsmakers die het systeem gebruiken om een beslissing te nemen. Bijvoorbeeld bij het aanvragen van een lening kan het systeem bepalen of de kandidaten aan de juiste voorwaarden voldoen. De beleidsmakers verwachten een uitleg die hun beslissing onderbouwt bij de kandidaten. Ten derde moeten regelgevende instanties zeker zijn dat de beslissingen gemaakt door het systeem veilig en eerlijk zijn. Ten slotte verwachten de eindgebruikers een uitleg voor de beslissing. De kandidaten van een lening willen bijvoorbeeld een duidelijke uitleg waarom ze een lening niet krijgen en willen meer horen dan “ons systeem zegt dat ik jullie de lening niet mag geven”. Er is een verschil tussen de uitleg voor beleidsmakers en eindgebruikers: de eerste hebben nood aan een technische uitleg en de tweede aan een meer laagdrempelige uitleg.

Metriecken voor XAI

Als gebruikers een AI-systeem gebruiken met verklaringen, kunnen ze geëvalueerd worden door gebruik te maken van metriecken, zoals correctheid van de uitleg, tevredenheid over de verklaringen, begrijpen van het AI-systeem, hoe nieuwsgierigheid het zoeken naar verklaringen motiveert, het vertrouwen van de gebruiker en de prestaties van het systeem [23]. Om de waargenomen kwaliteit van aanbevelingssystemen te evalueren, wordt vaak ResQue gebruikt; een vragenlijst met acht hoofdcategorieën die bestaan uit verschillende componenten [47]. Zo is er de categorie ‘gebruiksgemak’, die bestaat uit initieel leren, voorkeurspeiling, voorkeursherziening en besluitvorming.

Hierna wordt dieper ingegaan op een aantal metriecken, namelijk vertrouwen, controle en transparantie, interactie-adequatheid en waargenomen gebruiksgemak.

Vertrouwen. Niet elke gebruiker neemt zomaar aan dat de keuze van het systeem juist is, ze verwachten een rechtvaardiging voor de keuze [23]. Als het systeem erin

slaagt om de keuze te rechtvaardigen, door bijvoorbeeld transparantie, dan zal het een positief effect hebben op het vertrouwen van de gebruiker. Het vertrouwen van een gebruiker kan veranderen door het systeem te gebruiken en er zijn verschillende trajecten mogelijk. Gebruikers kunnen initieel veel vertrouwen hebben in AI-systemen, maar als ze tijdens het gebruik merken dat het systeem niet altijd de meest optimale keuze maakt, dan zal het vertrouwen dalen. In de literatuur worden verschillende schalen besproken om vertrouwen te evalueren en de lijst van Hoffman et al [23] staat in Tabel 2.5.

TABEL 2.5: Schalen voor vertrouwen [23].

Schaal	Jaar	Factoren
Cahour-Forzy	2009	Betrouwbaarheid, voorspelbaarheid en doeltreffendheid.
Jian et al.	2000	Trouw, loyaliteit, betrouwbaarheid, veiligheid, integriteit en vertrouwelijkheid.
Madsen-Gregor	2000	Betrouwbaarheid, technische bekwaamheid, begrijpelijkheid, geloof en persoonlijke verbondenheid.
Merritt	2011	Geloof, vertrouwen en betrouwbaarheid.
Schaefer	2013	Machine prestatie en groepswork (ontwikkeld voor mens-robot samenwerking).
Singh et al.	1993	Vertrouwen na ervaring, afhankelijkheid, initieel vertrouwen, veiligheid en zelfgenoegzaamheid.
Wang et al.	2009	Volledig gebaseerd op de schaal van Jian et al.

Ook Kato [26] gebruikte tijdens zijn onderzoek een schaal om vertrouwen te evalueren en onderscheidt twee types: eendimensionaal vertrouwen dat expliciet naar het vertrouwen vraagt, en meerdimensionaal vertrouwen dat onderverdeeld wordt in de componenten geloofsovertuiging (bekwaamheid, welwillendheid en integriteit), intentie om terug te keren en transparantie.

Hoewel er veel schalen bestaan, bestaat er twijfel over hun geldigheid, omdat vertrouwen moeilijk te meten/definiëren is [43]. Volgens Davis et al. [13] is vertrouwen een voordeel van een nuttig en betrouwbaar systeem. Vertrouwen gebruiken als een metriek voor een verklaring van het systeem is problematisch en kan daarnaast leiden tot kunstmatig opgeblazen niveaus van vertrouwen wat ten nadele is van de gebruiker. Het kan beter zijn om een metriek voor nut te ontwikkelen, dat onderzoekers aanzet om een bredere context van verklaringen te bekijken.

Controle en transparantie. Gebruikers die controle krijgen van een systeem kunnen hun voorkeuren herzien, aanbevelingen aanpassen en nieuwe aanbevelingen vragen [47]. Als ze geen controle krijgen, kunnen ze het systeem niet tegenhouden om telkens opnieuw slechte aanbevelingen te leveren en zo geraken ze teleurgesteld. Een transparant systeem laat daarnaast toe om de werking van het systeem te begrijpen, bijvoorbeeld via een verklaringsinterface [47].

Interactie-adequaatheid. Aanbevelingssysteem kunnen drie interactiemechanismen hebben, namelijk initiële voorkeursuitlokking, voorkeursherziening en verklaring

van de resultaten [47]. De integratie van die elementen in het aanbevelingssysteem heeft een grote impact op de perceptie van de gebruiker over het systeem.

Waargenomen gebruiksgemak. Waargenomen gebruiksgemak meet de mogelijkheid van de gebruiker om snel en correct taken uit te voeren [47]. Zoals eerder vermeld bestaat de metriek uit de componenten initieel leren, voorkeurspeiling, voorkeursherziening en besluitvorming.

2.3.3 XAI voor digitale leeromgevingen

Digitale systemen die zich aanpassen aan hun leerling hebben een model van elke leerling met een voorstelling van de capaciteiten en het gedrag gebaseerd op hun prestaties en interacties met het systeem [2]. Vroeger waren de modellen verborgen en wist de leerling niet welke informatie het systeem juist gebruikte [7]. Tegenwoordig stellen steeds meer systemen, namelijk *open learner models* (OLMs), de modellen open voor hun gebruikers, omdat zo de accuraatheid van het model verhoogt, zelfreflectie gepromoot wordt, leerlingen meer controle krijgen over hun leerproces en het vertrouwen in het systeem verhoogt [7].

Er zijn al verschillende leertechnologieën aangevuld met OLMs, waaronder *learning dashboards*, intelligente leersystemen en leersystemen die zich aanpassen om leerlingen en leerkrachten te helpen monitoren, reflecteren en reguleren [1].

Om aanbevelingen te verantwoorden kunnen educatieve aanbevelingssystemen OLMs gebruiken. Abdi et al. [2] vonden dat een OLM toevoegen aan een educatief aanbevelingssysteem studenten meer inzicht geeft in hun leerproces, wat een positief effect heeft op het engagement van de leerling en de perceptie van de effectiviteit van het aanbevelingssysteem. Het is wel belangrijk om een aantal elementen in het oog te houden. Het systeem moet voldoende informatie aanbieden zodat de leerling het OLM begrijpt en weet hoe het opgesteld wordt. Daarnaast moeten leerlingen kunnen aangeven wanneer ze het niet eens zijn met het OLM en moet de omgeving waarin ze het systeem gebruiken veilig zijn, ze moeten het gevoel hebben dat een fout niet onmiddellijk gevolgen heeft voor het model. OLMs kunnen echter ook nadelen met zich meebrengen. Ten eerste kan het de duidelijkheid van het systeem verlagen als de gebruiker het OLM niet begrijpt. Ten tweede kan het een gevoel van oneerlijkheid oproepen bij de gebruiken, wanneer hij/zij het niet eens is met het OLM. Tot slot kan het er ook voor zorgen dat het engagement van de leerling verlaagt uit schrik voor een negatief gevolg op het model.

2.4 Feedback in digitale leeromgevingen

Bij educatieve aanbevelingssystemen probeert het systeem de best mogelijke aanbevelingen te leveren voor de gebruiker. Het is mogelijk om hier *Interactive Machine Learning* (IML) aan te koppelen, waarbij de hulp van gebruikers ingeschakeld wordt om het systeem te verbeteren [20, 55]. Op deze manier leert het systeem niet alleen via de data die de gebruikers achterlaten, maar ook via feedback van de gebruikers.

Nochtans hebben de huidige digitale leeromgevingen vaak een feedbackmechanisme met als focus feedback geven aan de gebruiker om prestaties te verbeteren en veel minder vaak een feedbackmechanisme om feedback te geven als gebruiker.

2.4.1 Feedback krijgen als gebruiker

Leerlingen hebben nood aan verschillende soorten feedback om hun prestaties te verbeteren. Volgens Schumacher et al. [49] en Sedrakyan et al. [51] is het belangrijk om de feedback zo snel mogelijk te geven na het maken van een oefening. Hoe sneller, hoe sterker de vaardigheid om aan zelfregulatie te doen versterkt wordt.

Sedrakyan et al. [51] leggen enerzijds uit dat feedback procesgericht moet zijn. Vaak wordt er gezegd hoe de leerling het doet en niet hoe de leerling het beter kan doen. Feedback kan het proces vergemakkelijken door uit te leggen wat een goede prestatie is, zelfregulatie te vergemakkelijken en mogelijkheden aan te reiken om dichterbij een goede prestatie te komen.

Ze maken anderzijds het onderscheid tussen cognitieve en gedragsfeedback en beide zijn proces georiënteerd [51]. Bij het eerste wordt het leerproces gezien als een stap-voor-stap kennis constructie proces. Het helpt de leerling om te reflecteren over de kwaliteiten van zijn of haar probleem-oplossingsproces. Er zijn drie verschillende soorten cognitieve feedback. Ten eerste geeft corrigerende feedback de leerling informatie over de adequaatheid van het werk. Ten tweede stimuleert epistemische feedback uitleg en verklaringen vanuit de leerling zelf. Ten derde geeft suggestieve feedback advies aan de leerling voor verdere stappen. Volgens Admiraal et al. [4] moet feedback het leren van de leerling begeleiden door te vertellen wat nodig is om de prestatie te verbeteren. Figuur 2.3 toont de drie types die onderscheiden worden: feedback, *feed forward* en *feed up*.



FIGUUR 2.3: Drie types feedback [4, 31].

Verschillende onderzoekers zijn het er over eens dat wanneer begrijpbare feedback gegeven wordt aan leerlingen, het een bevorderend effect heeft [11, 28, 60].

2.4.2 Feedback geven als gebruiker

Naast het krijgen van feedback van een systeem kan de gebruiker ook feedback geven aan een systeem. Hierbinnen wordt een onderscheid gemaakt tussen expliciete en

impliciete feedback [32]. Bij expliciete feedback geven gebruikers bijvoorbeeld sterren aan films. Het is makkelijker te koppelen aan specifieke concepten, maar moeilijker om te verzamelen. Een voorbeeld van impliciete feedback is het aantal gekochte items van een klant. Het is makkelijker te verzamelen, maar de uitdaging bestaat erin de informatie te koppelen aan specifieke concepten. Expliciete en impliciete feedback zijn sterk ingeburgerd bij aanbevelingssystemen, maar in de context van een digitale leeromgeving zijn er minder applicaties. Een voorbeeld van het gebruik van expliciete feedback binnen digitale leeromgevingen is het vragen van feedback aan de leerling na het maken van een reeks oefeningen [35, 59]. De informatie wordt dan gebruikt door het systeem om betere aanbevelingen te leveren.

Eerder is het toevoegen van een OLM aan een educatief aanbevelingssysteem besproken. Het is mogelijk om de leerling het model te laten aanpassen door feedback te geven. Volgens Mabbott et al. [33] voelen veel leerlingen zich minder comfortabel wanneer ze volledige controle hebben over het model dan wanneer ze suggesties mogen doen, maar het systeem de eindbeslissing heeft.

Er moet opgemerkt worden dat het Dunning-Kruger effect kan optreden wanneer gebruikers feedback kunnen geven. Het effect wordt door Dunning et al. [16] gedefinieerd als volgt: “Zij die incompetent zijn, bij gebrek aan een betere term, hebben weinig inzicht in hun onbekwaamheid”. Het kan ook anders verwoord worden als: mensen die weinig kennis hebben over een bepaald onderwerp zouden niet in staat zijn om hun eigen zwakheden te herkennen, ook niet wanneer ze ernstig zijn. Het Dunning-Kruger effect kan dus bijvoorbeeld een rol spelen wanneer leerlingen hun eigen niveau moeten inschatten.

2.5 Gerelateerd werk

Er zijn al heel wat aanbevelingssystemen en digitale leeromgevingen ontwikkeld voor onderzoek. Wiski, ontwikkeld door Ooge [41], is een digitale leeromgeving waarbij de focus op *gamification* ligt en leerlingen uit het secundair onderwijs de doelgroep zijn. Ook RiPPLE [2] heeft een educatieve context en het is tegelijkertijd ook een aanbevelingssysteem. Het genereert een model van de leerling op basis van zijn/haar leeractiviteiten op het platform en bepaalt dan het niveau van de leerling met het Elo-classificatiesysteem. Op basis van het model en het niveau worden daarna nieuwe leeractiviteiten aanbevolen en bovendien heeft RiPPLE een OLM ingebouwd, waardoor leerlingen hun eigen model kunnen bekijken. Het volgende voorbeeld met een educatieve context is *Adaptive CSP applet* [10], dat ook interactie toelaat tussen het systeem en de gebruiker. Het helpt leerlingen via een interactieve simulatie om een algoritme te leren voor *constraint satisfaction problems*, met AI-gestuurde hints die zijn aangepast aan het voorspelde niveau van de leerling. Het systeem is later uitgebreid met een extra functionaliteit om de gebruiker verklaringen te geven. Eveneens hebben Adaptieve hypermedia [59] een educatieve context, ze zijn samengesteld uit een grote hoeveelheid leermateriaal gepresenteerd door hypertexten en willen de leerlingen niet begeleiden bij het oplossen van een probleem. De inschatting van het niveau van de leerling gebeurt met een algoritme gebaseerd op

het Elo-classificatiesysteem en dat van het cursusmateriaal wordt ook deels bepaald door de feedback van de leerling door het beantwoorden van vragen, zoals: “begrijp je de inhoud van het aanbevolen cursusmateriaal?”. Om te eindigen heeft ook ALEF (*Adaptive LEarning Framework*) [35] een educatieve context. Het produceert gepersonaliseerde aanbevelingen voor oefeningen en taken in een adaptieve digitale leeromgeving. Na elke poging om een oefening op te lossen, kan de leerling feedback geven via vier mogelijkheden: (1) De leerling heeft de oefening correct opgelost. (2) De leerling heeft de oefening correct opgelost met hulp of een hint. (3) De leerling heeft de oefening niet correct opgelost, maar begrijpt de oplossing. (4) De leerling heeft de oefening niet correct opgelost en begrijpt de oplossing niet. Twee extra voorbeelden die communiceren met de gebruiker zijn Glass-Box [52] en Robot traject planning [6]. Glass-Box communiceert met de gebruiker via een dialoog in natuurlijke taal waar verschillende “Waarom?” vragen gesteld kunnen worden om meer informatie te krijgen. Door interactie kan de gebruiker de verklaring dus uitbreiden en personaliseren. Robot traject planning genereert gepersonaliseerde verklaringen over de traject planning van een robot. De gebruiker heeft enerzijds de mogelijkheid om vragen te stellen over de verklaringen aan het systeem en anderzijds om wijzigingen aan te brengen aan de eigen voorkeuren om andere verklaringen te zien.

Hoofdstuk 3

Methodes

De literatuur toont twee leemtes. Ten eerste is er veel literatuur te vinden over feedback en digitale leeromgevingen, maar deze focussen grotendeels op het geven van feedback aan de leerling [11, 12, 28, 54, 60]. Feedback geven als gebruiker komt ook voor, maar weinig in de context van digitale leeromgevingen [35, 59]. Ten tweede zijn er weinig studies die focussen op het secundair onderwijs [4, 41]: de meeste focussen op studenten uit het hoger onderwijs [1, 2, 9, 10, 11, 24, 28, 29, 33, 35, 45, 49, 54, 60]. De pilootstudie met leerkrachten uit het secundair onderwijs en de leemtes hebben geleid tot volgende onderzoeksvragen.

Onderzoeksvraag 1. Hoe worden waargenomen gebruiksgemak en interactieadequaatheid beïnvloed door de toevoeging van de extra functionaliteiten?

Onderzoeksvraag 2. Hoe wordt het vertrouwen van leerlingen beïnvloed door het geven van controle over de aanbevelingen?

Onderzoeksvraag 3. Hoe wordt het vertrouwen van leerlingen beïnvloed door een uitleg over de impact van de controle van de leerling over de aanbevelingen?

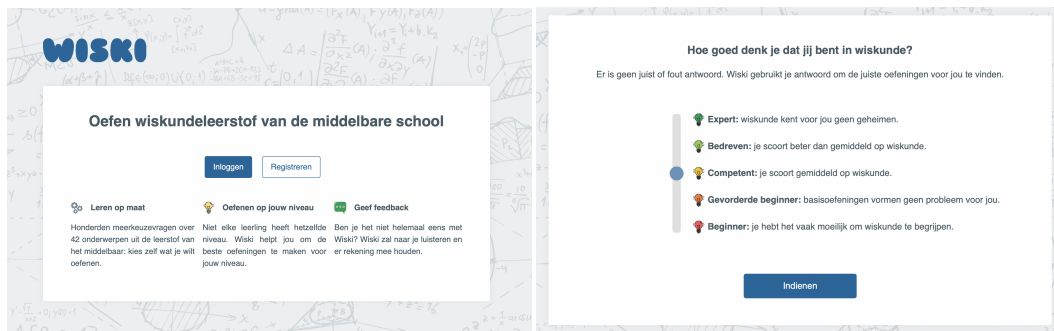
In dit hoofdstuk wordt de manier waarop de onderzoeksvragen beantwoord worden geschetst. Het uitgevoerde onderzoek is goedgekeurd door de Sociaal-Maatschappelijke Ethische Commissie (SMEC) (dossier: G-2022-4917).

3.1 Digitale leeromgeving Wiski

Het platform, Wiski, dat Ooge ontwikkelde in 2019 [41] is aangepast tot een platform waarbij oefeningen worden aanbevolen aan leerlingen op basis van hun niveau. Daarnaast krijgen de leerlingen controle over de aanbevelingen door na een reeks oefeningen aan te geven welke moeilijkheid van oefeningen ze willen voor de volgende reeks. Tot slot krijgen de leerlingen een uitleg van het platform over de impact van de controle. Tijdens het onderzoek worden er drie onderzoeksgroepen gebruikt, waarbij

3. METHODES

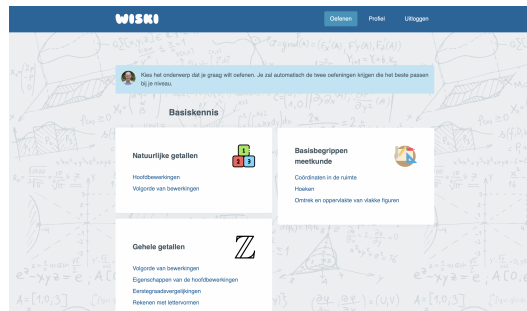
elke groep andere functionaliteiten heeft: (1) geen controle over de aanbevelingen (GEEN), (2) wel controle over de aanbevelingen (CONTROLE) en (3) controle en uitleg over de impact ervan (CONTROLE+IMPACT). Het overkoepelende doel is om te bepalen of de toevoeging van de functionaliteiten invloed heeft op waargenomen gebruiksgemak, interactie-adequaatheid en vertrouwen in de aanbevelingen. Leerlingen uit het secundair onderwijs die een doorstroomrichting, het vroegere ASO en TSO, volgen, vormen de doelgroep van het onderzoek. In deze thesis verwijst Wiski altijd naar het platform dat ontwikkeld is voor deze thesis, tenzij anders vermeld. Tijdens het onderzoek was het platform beschikbaar via <https://leen.wiski.be>. De belangrijkste pagina's van Wiski staan in Figuren 3.1 tot 3.5.



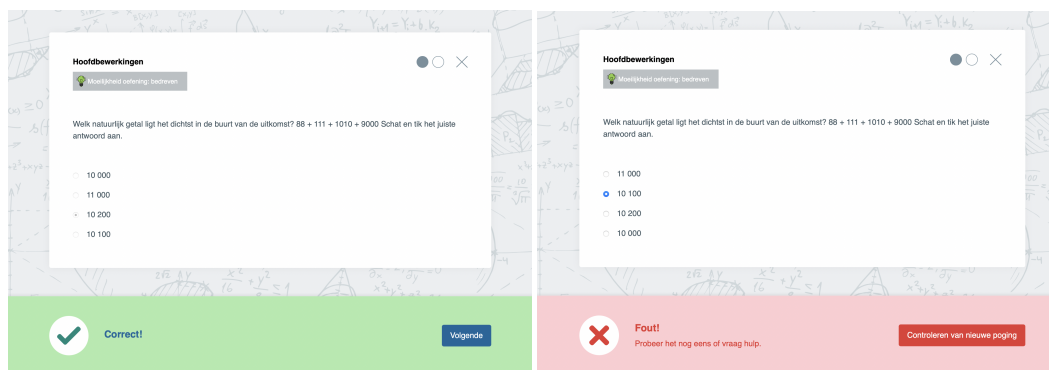
FIGUUR 3.1: Startpagina van Wiski (links). Pagina waarop gebruikers hun eigen startniveau kiezen net na de registratie (rechts).



FIGUUR 3.2: De pagina's geven uitleg over hoe Wiski werkt.



FIGUUR 3.3: Keuzepagina met onderwerpen van mogelijke oefeningen.



FIGUUR 3.4: Het scherm dat de gebruikers te zien krijgen wanneer ze een oefening juist (links) of fout (rechts) beantwoorden.



(A) Controle interface

(B) Impact van controle

FIGUUR 3.5: Controlepagina na elke reeks oefeningen (links). Bijhorende uitleg over de impact van de controle (rechts).

3.2 Pilootstudie

Om het doel van het onderzoek te bepalen is het belangrijk om te weten hoe leerkrachten zich verhouden tegenover digitale leeromgevingen, daarom is er gekozen

voor een pilootstudie. Hierbij werd het gevoel van de leerkrachten tegenover digitale leeromgevingen bevraagd. Er is gekozen voor een semi-gestructureerd interview waarbij er vragen werden gesteld aan de leerkracht om het gesprek te begeleiden [30]. De interviewer hield de tips van Leech [30] altijd in het achterhoofd, bijvoorbeeld het meermaals benadrukken van het belang om alles te zeggen wat de leerkracht denkt, ook als de leerkracht denkt dat de interviewer het al weet.

3.2.1 Deelnemers

De pilootstudie is uitgevoerd met vijf leerkrachten uit het secundair onderwijs. De deelnemers geven allemaal les in doorstroomrichtingen. De gesprekken duurden gemiddeld 44 minuten.

3.2.2 Opbouw

De pilootstudie bestond uit drie delen. Het eerste deel bevatte algemene vragen die de interviewer aan de leerkracht stelde (cfr. Appendix A.1), een deel van de vragen zijn gebaseerd op de werken van Admiraal et al. [4] en Michlík et al. [35]. Het doel bestond eruit een algemeen beeld te schetsen over de leerkracht en zijn of haar voorkeuren. Het tweede deel bestond uit het kiezen en sorteren van de drie belangrijkste items uit verschillende reeksen met bijvoorbeeld eigenschappen van een digitale leeromgeving en zijn gebaseerd op de werken van Millecamp et al. [36], Park et al. [45] en Pu et al. [47] (cfr. Appendix A.2). Het derde deel bestond uit voorbeelden van dashboards voor leerkrachten om leerlingen te monitoren.

3.3 Deelnemers gebruikersstudie

Zoals eerder aangegeven vormen leerlingen die een doorstroomrichting volgen aan de middelbare school de doelgroep van Wiski. Het verzamelen van testpersonen gebeurde op drie manieren. Ten eerste zijn negen scholen in het algemeen gecontacteerd. Hiervan hebben twee scholen gereageerd: één was geïnteresseerd, de andere ging het nog bekijken, maar heeft uiteindelijk niet deelgenomen. Ten tweede zijn 18 leerkrachten uit 11 scholen persoonlijk gecontacteerd via e-mail. In eerste instantie hebben zes leerkrachten toegestemd om deel te nemen, maar uiteindelijk hebben er nog drie afgezegd. Tot slot zijn 12 leerkrachten persoonlijk gecontacteerd via LinkedIn, geen enkele heeft gereageerd op het bericht. Aan de gebruikersstudie hebben uiteindelijk vier leerkrachten deelgenomen.

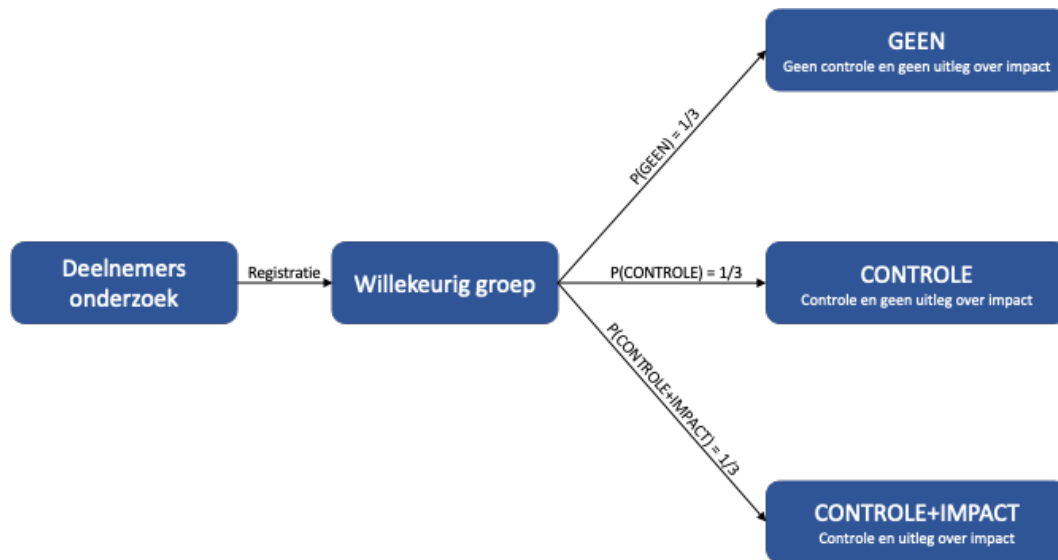
De geïnteresseerde leerkrachten kregen een informatiebrochure met de volgende onderdelen: het doel van het onderzoek, hoe het verloopt, wat er gebeurt met de verzamelde gegevens, voorzorgsmaatregelen voor Covid-19, de goedkeuring van SMEC en wat te doen bij vragen. Daarbovenop werd duidelijk gemaakt dat deelname een vrijwillige keuze is van de leerling en dat hij/zij niet onder druk mag gezet worden. Als een leerling niet wil deelnemen, moet de leerkracht een alternatief voorzien. De leerkrachten ontvingen ook een gelijkaardige brochure voor de leerlingen en hun ouders. De leerlingen moesten voor deelname een geïnformeerd toestemmingsformulier

onderteken. Als de leerling 15 jaar of jonger was, moesten ook de ouders het formulier ondertekenen. Al de voorgaande documenten kunnen bekeken worden in Appendix B.

3.4 Gerandomiseerd onderzoek met controlegroep

Om de onderzoeksvragen te beantwoorden is gekozen voor een gerandomiseerd onderzoek met controlegroep. Met een dergelijk onderzoek wordt het effect van een bepaald element op testpersonen bepaald, door middel van twee of meer onderzoeksgroepen. Voor het uitgevoerde onderzoek is het element waarover sprake is de gedeeltelijke controle van de gebruiker over de aanbevelingen en de uitleg over de impact ervan. Er is gekozen voor drie onderzoeksgroepen: (1) een groep die geen controle krijgt over de aanbevelingen (GEEN), (2) een groep die gedeeltelijke controle krijgt over de aanbevelingen (CONTROLE), maar geen uitleg krijgt over de impact ervan (3) een groep die gedeeltelijke controle krijgt over de aanbevelingen en een uitleg krijgt over de impact ervan (CONTROLE+IMPACT). De eerste groep is tijdens het onderzoek de controlegroep. Het doel van het gerandomiseerd onderzoek met controlegroep is om te bepalen of de controle en de uitleg over de impact ervan tot een gelijkaardig waargenomen gebruiksgemak, betere interactie-adequaatheid en meer vertrouwen in de aanbevelingen leidt dan wanneer de gebruiker geen controle krijgt.

Tijdens de registratie worden de gebruikers willekeurig onderverdeeld in de drie onderzoeksgroepen. Het is belangrijk om dit willekeurig te doen, zodat er geen selectie vertekening kan ontstaan. Daarnaast is het gebruikte platform identiek voor alle groepen op de controle en de uitleg over de impact ervan na. Op die manier kunnen geobserveerde verschillen hoogstwaarschijnlijk toegewezen worden aan de twee elementen. De werkwijze staat samengevat in Figuur 3.6.



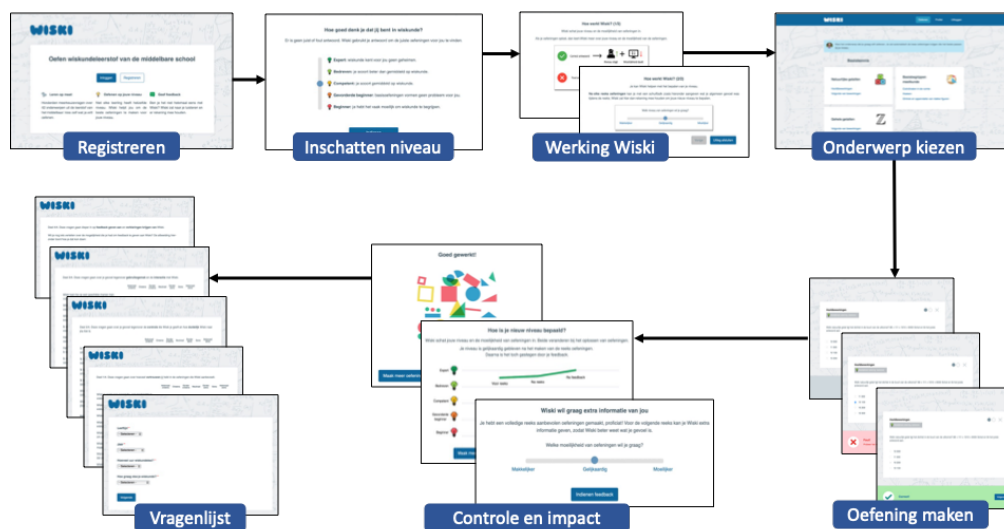
FIGUUR 3.6: Visualisatie gerandomiseerd onderzoek met controlegroep.

3. METHODES

Er moet opgemerkt worden dat de onderzoeken gebeurden tijdens wiskundelessen. Het was dus mogelijk dat leerlingen het scherm zagen van andere leerlingen, die behoorden tot een andere onderzoeksgroep, wat kon leiden tot verwarring en de eindresultaten kon beïnvloeden. Een alternatief was om iedereen uit dezelfde klasgroep toe te wijzen aan dezelfde onderzoeksgroep. Niettemin zorgt het voor selectie vertekening omdat de onderzoeksgroep dan afhangt van de school en de achtergrondkennis van de leerlingen. Daarom werd er gekozen om de leerlingen toch volledig willekeurig toe te wijzen aan de verschillende groepen.

3.4.1 Verloop van het onderzoek

Het uitgevoerde onderzoek bestond uit drie fasen en het verloop ervan staat afgebeeld in Figuur 3.7.



FIGUUR 3.7: Verloop van het onderzoek.

Tijdens de eerste fase, de registratie, moesten de deelnemers een account aanmaken en daarna hun eigen startniveau kiezen, de eerste reeks oefeningen werd dan bepaald op basis van dat niveau. Bovendien kregen ze een korte uitleg over hoe Wiski werkt, namelijk hoe het algoritme werkt en hoe de gebruiker controle krijgt.

Tijdens de tweede fase konden de deelnemers Wiski gebruiken. Er werden telkens reeksen van twee oefeningen aanbevolen en na elke reeks kregen **CONTROLE** en **CONTROLE+IMPACT** controle over de volgende reeks aanbevelingen en **CONTROLE+IMPACT** kreeg bovendien een uitleg over de impact van de controle. Er is gekozen voor reeksen van twee oefeningen omdat na de pilootstudie bleek dat leerkrachten de voorkeur gaven aan meerdere vragen, zodat een vergissing niet onmiddellijk afgestraft wordt. Meer dan twee vragen was niet mogelijk omdat de deelnemers voldoende keren de kans moesten krijgen om de volgende reeks aanbevelingen aan te passen en de eventueel bijhorende uitleg over de impact ervan te bekijken.

Tijdens de derde fase moesten de deelnemers een vragenlijst invullen, die automatisch kwam na het maken van drie reeksen en eventueel het uitoefenen van controle en het bekijken van de uitleg over de impact ervan. Alle deelnemers kregen de vragenlijst op hetzelfde moment, zodat ze allemaal evenveel ervaring hadden tijdens het invullen. Drie reeksen van twee oefeningen komt neer op het oplossen van zes oefeningen. Het onderzoek van Ooge [41] toonde aan dat gebruikers ongeveer 15 oefeningen oplosten per uur, maar ze moesten achteraf geen vragenlijst invullen, daarom lijkt zes oefeningen een veilige keuze. Na het invullen van de vragenlijst mochten de deelnemers Wiski vrij verder gebruiken. Als er nog andere klasgroepen gepland stonden om deel te nemen aan het onderzoek werd er gevraagd om enkel oefeningen te maken horende bij de eigen jaren, zodat de moeilijkheden van de oefeningen van andere jaren niet werden aangepast door een andere doelgroep.

3.5 Oefeningen aanbevelen

Zoals eerder besproken is Wiski een aanbevelingssysteem waaraan twee elementen toegevoegd zijn. Ten eerste krijgt de gebruiker gedeeltelijke controle over de aanbevelingen, na elke reeks oefeningen kan de gebruiker aangeven welke moeilijkheid de voorkeur krijgt voor de volgende reeks. Ten tweede geeft Wiski een uitleg over de impact van de controle.

3.5.1 Algoritme

De focus van het uitgevoerde onderzoek is niet het beste aanbevelingssysteem vinden, maar de invloed bepalen van de twee toegevoegde elementen. Er is daarom gekozen voor een bestaand algoritme met goede resultaten, namelijk het Elo-classificatiesysteem van Pelànek [46]. Een voordeel van het Elo-classificatiesysteem is de mogelijkheid om makkelijk veranderingen in niveau bij te houden. Om het algoritme te kunnen gebruiken, moeten de waarden van drie elementen gekozen worden: de initiële niveaus van de gebruikers, de initiële moeilijkheden van de oefeningen en de waarde van de K -parameter. Het Elo-classificatiesysteem is initieel ontwikkeld om een klassement op te stellen voor schaakspelers. De waarden die hieraan verbonden zijn, liggen ongeveer tussen 1000 en 2000 en dat bereik wordt ook gebruikt voor de niveaus van de gebruikers en de moeilijkheden van de oefeningen.

Initieel niveau gebruiker. Om het *cold start* probleem in verband met de niveaus van de gebruikers op te lossen, is er gekozen om de gebruikers zelf hun startniveau te laten kiezen. Het bereik tussen 1000 en 2000 is omgezet in een schuifbalk waarmee de gebruikers hun keuze kunnen maken. Er werden vijf niveaus (beginner, gevorderde beginner, competent, bedreven en expert) met een beschrijving toegevoegd, zodat het voor de gebruikers makkelijker is hun eigen niveau te bepalen (cfr. Figuur 3.1).

Initiële moeilijkheid oefeningen. Om het *cold start* probleem in verband met de moeilijkheid van de oefeningen op te lossen, is er gekozen om een expert in te

schakelen. De deelnemende leerkrachten konden voor deelname aan het onderzoek kiezen over welke onderwerpen hun leerlingen oefeningen moesten maken. Er werd dan gevraagd aan de leerkrachten om een moeilijkheid te bepalen voor elke oefening van dat onderwerp (cfr. Appendix E). Het bereik tussen 1000 en 2000 geeft te veel keuzes voor leerkrachten, aangezien ze zelf aangaven dat vijf keuzemogelijkheden voldoende was, konden ze de moeilijkheden aanduiden met de niveaus die de gebruikers ook te zien krijgen. Een aantal leerkrachten wilden graag meerdere onderwerpen aanbieden aan hun leerlingen en hebben de moeilijkheid bepaald van alle oefeningen horende bij de gekozen onderwerpen.

***K*-parameter.** Voor de waarde van de *K*-parameter moet een afweging gemaakt worden [46]: als de waarde van *K* te klein is, zal de schatting te traag convergeren; als de waarde te groot is, zal de schatting onstabiel zijn omdat het een te grote invloed geeft aan de laatste pogingen. Eerder werk in verband met modellen in een educatieve context gebruikte de waarde 0,4 en die waarde is ook gebruikt tijdens het uitgevoerde onderzoek [46].

3.5.2 Controle

Na elke reeks oefeningen krijgt de gebruiker gedeeltelijke controle over de volgende aanbevelingen via een schuifbalk waarachter waarden zitten tussen 0,90 en 1,10 (cfr. Figuur 3.5a). De gebruiker krijgt de exacte waarden niet te zien, maar er zijn wel labels toegevoegd aan de schuifbalk om te verduidelijken. De gekozen waarde wordt vermenigvuldigd met het huidige niveau en het resultaat van de vermenigvuldiging is het nieuwe niveau van de gebruiker. Tijdens het onderzoek krijgen CONTROLE en CONTROLE+IMPACT gedeeltelijke controle over de aanbevelingen.

3.5.3 Uitleg over impact van controle

Nadat gebruikers controle hebben gekregen over de volgende aanbevelingen, krijgen ze een uitleg over de impact ervan op hun niveau (cfr. Figuur 3.5b) en die bestaat uit drie delen: waarom, verantwoording en een visualisatie. Het eerste deel over *waarom* de gebruiker de uitleg krijgt, is telkens hetzelfde en beschrijft dat Wiski het niveau inschat van de gebruiker en de moeilijkheid van de oefeningen en dat beide veranderen door het oplossen van oefeningen. Het tweede deel is een *verantwoording* voor de verandering van het niveau van de gebruiker door te oefenen en de controle. Het derde deel is een *visualisatie* van de verantwoording om het extra te verduidelijken. Tijdens het onderzoek krijgt enkel CONTROLE+IMPACT de uitleg.

3.6 Vragenlijst gebruikt na het onderzoek

In deze paragraaf worden de keuzes die gemaakt zijn om tot de gebruikte vragenlijst te komen besproken. De vragenlijst, die volledig terug te vinden is in Appendix D, bestaat uit vijf thema's: demografie, vertrouwen, interactie-adequaatheid, waargenomen gebruiksgemak en verduidelijking over het krijgen van controle en het krijgen

van een uitleg over de impact ervan. De vragen uit het thema demografie worden gesteld om meer inzicht te krijgen in algemene informatie over de gebruiker. Hoe de laatste vier thema's behandeld worden door de literatuur werd eerder besproken in Paragraaf 2.3.2. Bijna alle vragen uit de vier thema's worden bevraagd via een 7-punt Likertschaal. De anderen zijn open vragen en daarvoor is antwoorden enkel verplicht voor het thema vertrouwen.

Het thema vertrouwen is verder onderverdeeld in verschillende componenten gebaseerd op het onderzoek van Kato [26]. Enerzijds is er multidimensionaal vertrouwen dat verder onderverdeeld is in geloofsovertuigingen, intentie om terug te keren en transparantie, anderzijds is er eendimensionaal vertrouwen. Geloofsovertuigingen wordt ook nog eens verder onderverdeeld in bekwaamheid, welwillendheid en integriteit. Het deel over transparantie bestond bij het onderzoek van Kato [26] uit één vraag, maar aangezien transparantie een belangrijk onderdeel is van het onderzoek werd beslist transparantie uit te breiden tot controle en transparantie. De beslissing is genomen op basis van de bespreking van controle en transparantie door Pu et al. [47]. De gebruikte vragen voor die component zijn overgenomen uit hun bespreking.

Naast vertrouwen is er gekozen om nog drie extra thema's te bevragen, zodat er een diepgaander onderzoek kan gebeuren. Alle drie zijn ze van cruciaal belang om te bepalen of de controle en de bijhorende uitleg over de impact ervan een positieve invloed hebben op het gevoel van de gebruiker. De vragen voor interactie-adequaatheid en waargenomen gebruiksgemak zijn overgenomen uit ResQue van Pu et al. [47]. Voor waargenomen gebruiksgemak zijn twee van de vier componenten besproken in Paragraaf 2.3.2 gekozen, namelijk voorkeurspeiling en voorkeursherziening, omdat de andere twee niet relevant zijn voor het gebruikte platform. Het laatste thema vraagt expliciet aan de gebruiker wat ze vinden van de gedeeltelijke controle over de aanbevelingen en een uitleg over de impact ervan. Voor GEEN en CONTROLE zijn de vragen geherformuleerd: GEEN krijgt geen controle en dus ook geen uitleg over de impact ervan, daarom wordt gevraagd of ze het graag zouden willen. Als verduidelijking zijn er afbeeldingen toegevoegd over hoe het verwerkt zit in het platform. CONTROLE krijgt wel controle, waardoor hier expliciet gevraagd wordt naar hun mening. Ze krijgen echter geen uitleg over impact ervan, dus hier wordt gevraagd of ze dat graag willen met een afbeelding als verduidelijking.

3.7 Statistische analyse

Er werd zowel een statistische analyse uitgevoerd als een kwalitatieve bespreking. De meeste vragen uit de vragenlijst worden beantwoord via een Likertschaal, daarom wordt de statistische analyse een combinatie van de Mann-Whitney U toets en de ongepaarde t-toets afhankelijk van de situatie [58]. Als een metriek geanalyseerd wordt met maar één vraag, dan wordt de Mann-Whitney U toets gebruikt, omdat er telkens twee groepen vergeleken worden. Als een metriek uit meerdere vragen bestaat, worden de condities voor de ongepaarde t-toets bekeken: de groepen moeten onafhankelijk zijn, beide groepen moeten normaal verdeeld zijn en de varianties van de groepen moeten gelijk zijn [56]. De groepen zijn voor alle uitgevoerde vergelijkingen

onafhankelijk omdat ze andere functionaliteiten omvatten. Of beide groepen normaal verdeeld zijn, wordt bepaald aan de hand van de Shapiro-Wilk toets. Als ze niet normaal verdeeld zijn, is de Mann-Whitney U toets uitgevoerd, anders werd bepaald of de varianties van de groepen gelijk zijn aan de hand van de F-toets [50, 56]. Als ze gelijk zijn, wordt de Student t-toets uitgevoerd en anders de Welch t-toets [56]. Om de statistische analyse uit te voeren is er gebruik gemaakt van R.

Waargenomen gebruiksgemak. Om de resultaten tussen de verschillende groepen te vergelijken werd per twee een analyse uitgevoerd. Aangezien er telkens verwacht werd dat er geen verschil was tussen de groepen op vlak van waargenomen gebruiksgemak, was een tweezijdige toets geschikt. De uitgevoerde toetsen per vergelijking voor waargenomen gebruiksgemak staan in Tabel 3.1. De opbouw van de hypothesen voor elke toets zijn telkens hetzelfde: de nulhypothese stelt dat er geen verschil is tussen de gemiddelden van de twee groepen, de alternatieve hypothese stelt dat er wel een verschil is tussen de gemiddelden van de twee groepen.

Hypothese H_0 : $gem_X = gem_Y$

Hypothese H_1 : $gem_X \neq gem_Y$

met (X,Y) uit {(GEEN, CONTROLE), (GEEN, CONTROLE+IMPACT), (CONTROLE, CONTROLE+IMPACT)}

TABEL 3.1: Uitgevoerde toets per vergelijking voor waargenomen gebruiksgemak.

(A) GEEN vs. CONTROLE	
	Toets
Voorkeurspeiling	Mann-Whitney U
Voorkeursherziening	Student t-test
Waargenomen gebruiksgemak	Mann-Whitney U
(B) GEEN vs. CONTROLE+IMPACT	
	Toets
Voorkeurspeiling	Mann-Whitney U
Voorkeursherziening	Student t-test
Waargenomen gebruiksgemak	Mann-Whitney U
(C) CONTROLE vs. CONTROLE+IMPACT	
	Toets
Voorkeurspeiling	Mann-Whitney U
Voorkeursherziening	Student t-test
Waargenomen gebruiksgemak	Student t-test

Interactie-adequaatheid. Om de resultaten tussen de verschillende groepen te vergelijken werd per twee een analyse uitgevoerd. Aangezien er telkens verwacht werd dat één groep een betere interactie-adequaatheid had dan de andere, was een

eenzijdige toets geschikt. De uitgevoerde toetsen per vergelijking voor interactieadequatheid staan in Tabel 3.2. De opbouw van de hypothesen voor elke toets zijn telkens hetzelfde: de nulhypothese stelt dat er geen verschil is tussen de gemiddelden van de twee groepen, de alternatieve hypothese stelt dat de groep met meer functionaliteiten een hoger gemiddelde heeft:

Hypothese H_0 : $gem_X = gem_Y$

Hypothese H_1 : $gem_X < gem_Y$

met (X,Y) uit {(GEEN, CONTROLE), (GEEN, CONTROLE+IMPACT), (CONTROLE, CONTROLE+IMPACT)}

TABEL 3.2: Uitgevoerde toets per vergelijking voor interactieadequatheid.

	Toets
GEEN vs. CONTROLE	Student t-toets
GEEN vs. CONTROLE+IMPACT	Student t-toets
CONTROLE vs. CONTROLE+IMPACT	Welch t-toets

Vertrouwen in aanbevelingen. Om de resultaten tussen de verschillende groepen te vergelijken werd per twee een analyse uitgevoerd. Aangezien er telkens verwacht werd dat één groep meer vertrouwen had in de aanbevelingen dan de andere, was een eenzijdige toets geschikt. De uitgevoerde toetsen per vergelijking voor vertrouwen in aanbevelingen staan in Tabel 3.3. De opbouw van de hypothesen voor elke toets zijn telkens hetzelfde: de nulhypothese stelt dat er geen verschil is tussen de gemiddelden van de twee groepen, de alternatieve hypothese stelt dat de groep met meer functionaliteiten een hoger gemiddelde heeft.

Hypothese H_0 : $gem_X = gem_Y$

Hypothese H_1 : $gem_X < gem_Y$

met (X,Y) uit {(GEEN, CONTROLE), (GEEN, CONTROLE+IMPACT), (CONTROLE, CONTROLE+IMPACT)}

TABEL 3.3: Uitgevoerde toets per vergelijking voor vertrouwen in aanbevelingen.

(A) GEEN vs. CONTROLE

	Toets
Bekwaamheid	Student t-toets
Welwillendheid	Mann-Whitney U
Integriteit	Student t-toets
Intentie om terug te keren	Student t-toets
Controle en transparantie	Student t-toets
Eendimensionaal vertrouwen	Mann-Whitney U
Geloofsovertuiging	Student t-toets
Multidimensionaal vertrouwen	Student t-toets

3. METHODES

(B) GEEN vs. CONTROLE+IMPACT

	Toets
Bekwaamheid	Student t-toets
Welwillendheid	Mann-Whitney U
Integriteit	Student t-toets
Intentie om terug te keren	Mann-Whitney U
Controle en transparantie	Student t-toets
Eendimensionaal vertrouwen	Mann-Whitney U
Geloofsovertuiging	Student t-toets
Multidimensionaal vertrouwen	Student t-toets

(C) CONTROLE vs. CONTROLE+IMPACT

	Toets
Bekwaamheid	Student t-toets
Welwillendheid	Student t-toets
Integriteit	Student t-toets
Intentie om terug te keren	Mann-Whitney U
Controle en transparantie	Student t-toets
Eendimensionaal vertrouwen	Mann-Whitney U
Geloofsovertuiging	Student t-toets
Multidimensionaal vertrouwen	Student t-toets

Hoofdstuk 4

Ontwikkeling

De digitale leeromgeving Wiski werd ontwikkeld via een iteratief gebruikersgestuurd ontwerp, waarbij de eindgebruikers meermaals bij het productieproces betrokken werden om problemen snel op te sporen (cfr. Figuur 4.1).



FIGUUR 4.1: Ontwikkelingsproces van Wiski.

Na de pilootstudie werden twee iteraties met *think-aloud* studies uitgevoerd om tot het finaal *proof of concept* te komen. Tijdens een *think-aloud* studie doorlopen de testgebruikers het prototype via opdrachten. Tegelijkertijd vertellen ze al hun gedachten luidop en het is niet de bedoeling dat de opdrachtgever tussenkomt bij fouten of moeilijkheden, zodat het mogelijk is om problemen snel op te sporen. Voor de *think-aloud* studies werden prototypes gemaakt. Voor het eerste prototype werd gekozen voor een *low-fidelity* prototype, wat een functioneel uitgedacht systeem is waarbij weinig rekening wordt gehouden met de vorm. Om snel een functioneel prototype te bekomen, werd Microsoft PowerPoint gebruikt. Voor het tweede prototype werd het *low-fidelity* prototype omgevormd tot een digitaal prototype, het *high-fidelity* prototype, en de problemen werden aangepast. Het finaal *proof of concept* is de uiteindelijke versie van Wiski dat gebruikt werd tijdens het onderzoek.

4.1 Pilootstudie

In deze paragraaf worden de belangrijkste resultaten uit de pilootstudie besproken. De belangrijkste kenmerken van het ontwikkelde platform, Wiski, zijn daarop gebaseerd. De volledige resultaten zijn terug te vinden in Appendix A.3. Eén van de deelnemers gaf aan geen digitale leeromgevingen te gebruiken, waardoor hij het niet nuttig vond om de hele pilootstudie te doorlopen. De resultaten die hier besproken worden zijn dus afkomstig van de overige vier leerkrachten.

De leerkrachten gaven allemaal aan dat Smartschool en digitale bordboeken de digitale platformen zijn die ze vooral gebruiken. Daarnaast worden er ook andere platformen gebruikt, maar steeds maar door één leerkracht (bv. Code.org en Usolv-it).

Verder gaven drie leerkrachten aan dat meerkeuzevragen hun voorkeur niet krijgen, maar dat ze begrijpen waarom ervoor gekozen wordt. De redenen waarom ze het liever niet gebruiken zijn uiteenlopend. Een eerste leerkracht zei: *“Het denkproces is belangrijk, zeker bij diegenen die remediëring nodig hebben.”* Voor de tweede leerkracht was de mogelijkheid om te gokken de grootste reden en de laatste gaf aan het liever niet te zien, maar ze ziet zelf geen andere mogelijkheid. De keuze voor meerkeuzevragen is ook de reden waarom geen enkele leerkracht dergelijk systeem zou gebruiken voor evaluatie, maar wel voor remediëring. De leerlingen zouden nooit verplicht worden om een digitale leeromgeving te gebruiken en in de les is er geen tijd voor. Er werd ook aangegeven dat de leerkrachten graag zelf oefeningen zouden toevoegen, zodat ze de remediëring beter kunnen sturen.

Van oefeningen aanbevelen aan de leerlingen zijn alle leerkrachten fan. Er zijn drie leerkrachten die enerzijds de aanbevelingen graag zouden sturen en anderzijds ook leerlingen zouden toelaten om de aanbevelingen te sturen. Al de leerkrachten verkiezen een systeem waarbij meerdere oefeningen worden aanbevolen voordat het systeem nieuwe oefeningen aanbeveelt, onder het motto *“Iedereen kan al eens een fout maken, moet de leerling dan onmiddellijk afgestraft worden?”* Ze willen wel liefst dat de aanbevolen oefeningen verplicht zijn. Over het tonen van de juiste oplossing van een oefening zijn drie leerkrachten het eens. Als je de oplossing toont, moet het een uitgewerkte versie zijn, zodat de leerling zijn of haar denkfout kan achterhalen. Voor de vierde leerkracht is enkel de juiste oplossing ook voldoende.

Uit het sorteren van de reeksen kwamen de volgende belangrijke eigenschappen voor het systeem: gebruiksvriendelijk, divers, efficiënt, transparant en instelbaar.

De kernelementen van de dashboards die de leerkrachten te zien kregen en het aantal leerkrachten die dat belangrijk vonden, worden samengevat in Tabel 4.1.

TABEL 4.1: Aantal leerkrachten die een bepaald kernelement voor dashboards belangrijk vonden.

Kernelement	#Leerkrachten
Notificaties voor leerkrachten met belangrijke inzichten.	4
Naast data voor specifieke leerlingen ook data tonen voor specifieke oefeningen.	3
Leerlingen clusteren.	3
Hoe meer details, hoe beter.	3
Groei-curve van individuele leerling.	1
Klassen vergelijken.	0

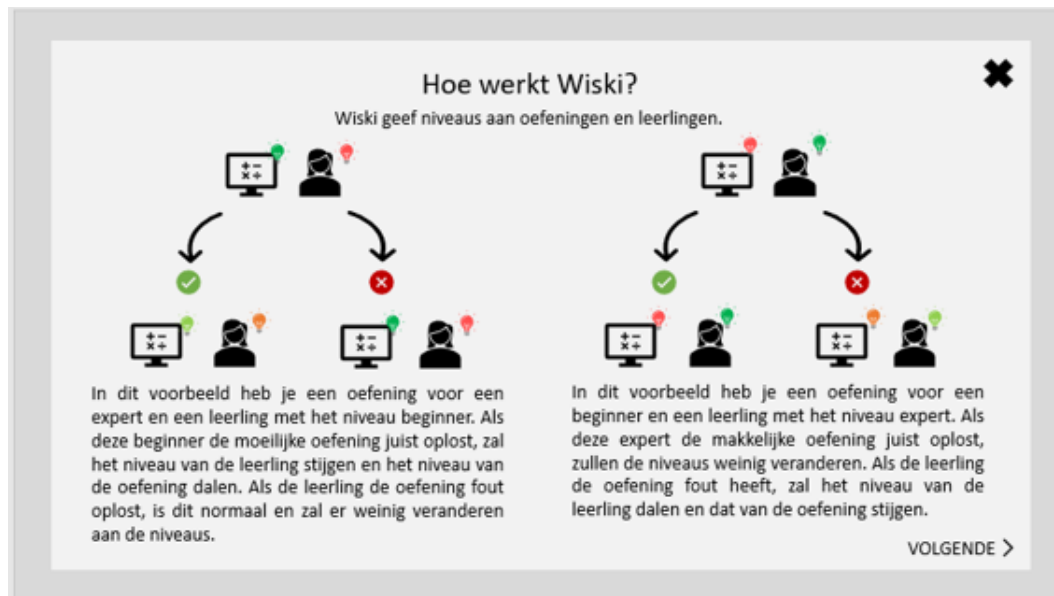
Na de pilootstudie werd beslist om te focussen op een aanbevelingssysteem met drie kenmerken. Ten eerste kiest de leerkracht, domein expert, de initiële moeilijkheid van de oefeningen. Ten tweede krijgt de leerling gedeeltelijke controle over de aanbevolen oefeningen. Ten slotte geeft het systeem een uitleg aan de leerlingen over wat de impact is van de controle.

4.2 *Low-fidelity* prototype

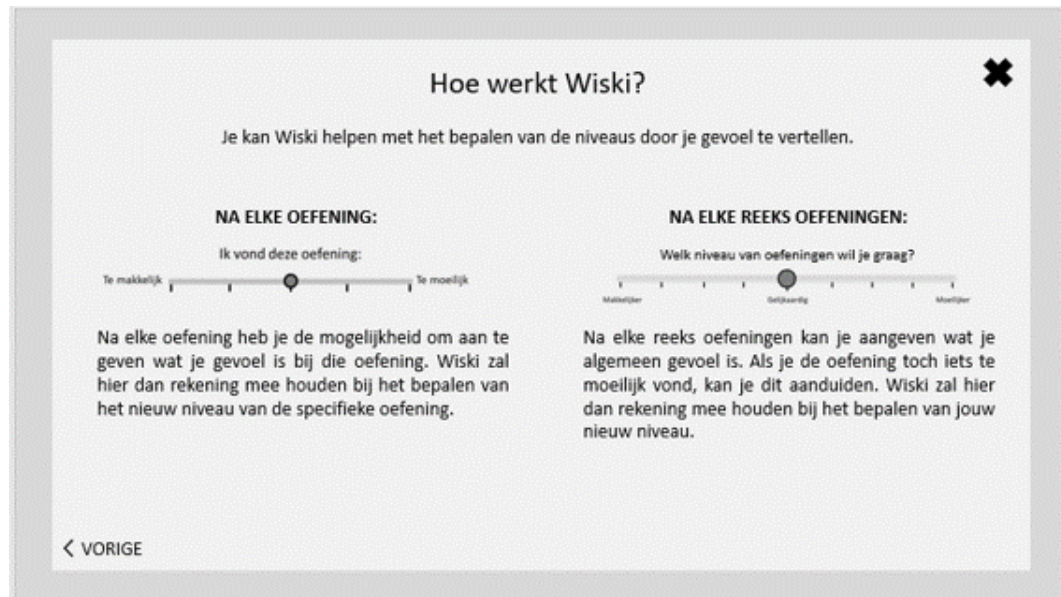
Het doel van het *low-fidelity* prototype was om vier sleutelementen te testen bij potentiële gebruikers, namelijk: gebruiksvriendelijkheid, visualisatie voor uitleg over impact van controle, uitleg over werking van systeem en de vragenlijst voor het uiteindelijke onderzoek. Figuren 4.2 tot 4.10 tonen de belangrijkste elementen uit het prototype.

Zeven potentiële gebruikers namen deel aan de eerste *think-aloud* studie. Vijf gebruikers studeerden Wiskunde-Wetenschappen in het vijfde middelbaar en waren 16 jaar. De overige twee studeerden Moderne Talen-Wetenschappen-Sport in het tweede middelbaar en waren 13 jaar. Wegens tijdsgebrek heeft één gebruiker uit het tweede middelbaar de vragenlijst niet doorlopen. De resultaten hiervan werden dus gebaseerd op zes gebruikers.

De vragen die gebruikt werden tijdens de *think-aloud* studie staan in Appendix C. De vragen waren opgedeeld in twee delen. In het eerste deel werd gefocust op een vlotte navigatie doorheen het prototype, waardoor maar één van de drie mogelijke visualisaties getoond werd, namelijk de lijngrafiek. Tijdens het tweede deel was er een algemene bespreking waarbij de andere twee visualisaties getoond werden. Er werd ook gepeild naar welke visualisatie de voorkeur kreeg en de resultaten waren niet eenduidig: zowel de lijngrafiek als het histogram werden drie keer gekozen, de bar-grafiek werd slechts één keer gekozen.



FIGUUR 4.2: Uitleg over werking Wiski (algoritme).



FIGUUR 4.3: Uitleg over werking Wiski (controle leerling).



FIGUUR 4.4: Pagina waar de leerling een start-level kiest voor het gekozen onderwerp.

Hoofdbewerkingen

Welk natuurlijk getal ligt het dichtst in de buurt van de uitkomst? $5210 - 2180$ Schat en tik het juiste antwoord aan.

☒ 30
☐ 300
☐ 7300
☐ 3000

FOUT Ik vond deze oefening: Te makkelijk Te moeilijk **VERDER**

FIGUUR 4.5: Pagina na het fout oplossen van een oefening. De leerling krijgt controle over de moeilijkheid van de oefening.

Hoofdbewerkingen

Welk natuurlijk getal ligt het dichtst in de buurt van de uitkomst? $5210 - 2180$ Schat en tik het juiste antwoord aan.

☐ 30
☐ 300
☐ 7300
☒ 3000

CORRECT Ik vond deze oefening: Te makkelijk Te moeilijk **VERDER**

FIGUUR 4.6: Pagina na het juist oplossen van een oefening. De leerling krijgt controle over de moeilijkheid van de oefening.

Wiski wil graag extra informatie van jou

Je hebt een volledige reeks aanbevolen oefeningen gemaakt, proficiat! Voor de volgende reeks kan je Wiski extra informatie geven, zodat Wiski beter weet wat je gevoel is.

Welk niveau van oefeningen wil je graag?



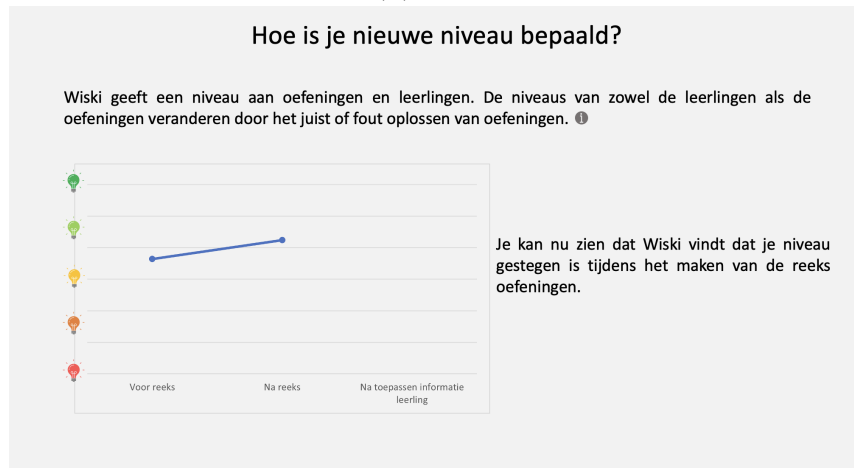
Makkelijker Gelijkwaardig Moeilijker

NIEUWE OEFENINGEN ANNULEREN

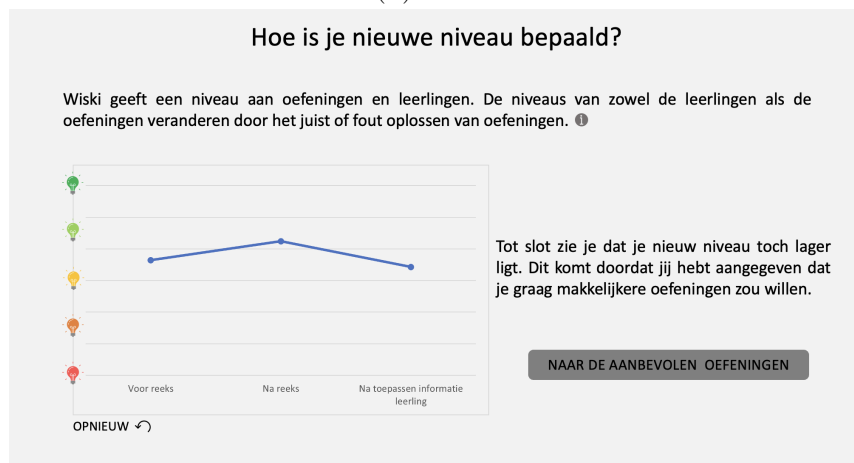
FIGUUR 4.7: Pagina waar de leerling kan aangeven welk niveau gewenst is voor de volgende reeks oefeningen.



(A) Deel 1



(B) Deel 2

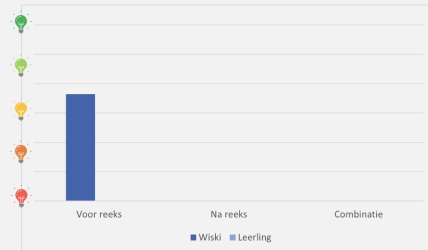


(c) Deel 3

FIGUUR 4.8: Uitleg met animatie over de impact van de controle (lijngrafiek).

Hoe is je nieuwe niveau bepaald?

Wiski geeft een niveau aan oefeningen en leerlingen. De niveaus van zowel de leerlingen als de oefeningen veranderen door het juist of fout oplossen van oefeningen. ①

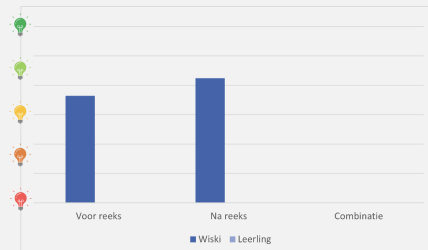


De figuur hiernaast zal tonen hoe jouw niveau veranderd is door het maken van de vorige reeks oefeningen. Op dit moment is je niveau van voor het maken van de reeks oefeningen zichtbaar.

(A) Deel 1

Hoe is je nieuwe niveau bepaald?

Wiski geeft een niveau aan oefeningen en leerlingen. De niveaus van zowel de leerlingen als de oefeningen veranderen door het juist of fout oplossen van oefeningen. ①

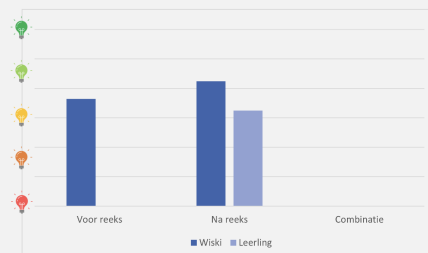


Je kan nu zien dat Wiski vindt dat je niveau gestegen is tijdens het maken van de reeks oefeningen.

(B) Deel 2

Hoe is je nieuwe niveau bepaald?

Wiski geeft een niveau aan oefeningen en leerlingen. De niveaus van zowel de leerlingen als de oefeningen veranderen door het juist of fout oplossen van oefeningen. ①



Na het maken van de reeks oefeningen heb je zelf aangegeven wat je gevoel is. De lichtblauwe balk toont hoe jij je niveau hebt ingeschat.

(C) Deel 3



(D) Deel 4

FIGUUR 4.9: Uitleg met animatie over de impact van de controle (histogram).



(A) Deel 1

Hoe is je nieuwe niveau bepaald?

Wiski geeft een niveau aan oefeningen en leerlingen. De niveaus van zowel de leerlingen als de oefeningen veranderen door het juist of fout oplossen van oefeningen. ⓘ



Op dit moment zie je hoe jouw niveau veranderd is tijdens het maken van de reeks oefeningen.

(B) Deel 2

Hoe is je nieuwe niveau bepaald?

Wiski geeft een niveau aan oefeningen en leerlingen. De niveaus van zowel de leerlingen als de oefeningen veranderen door het juist of fout oplossen van oefeningen. ⓘ



Na het maken van de reeks oefeningen heb je Wiski kunnen vertellen wat jouw gevoel was. Het gearceerde deel toont de verandering die jouw gevoel teweeg heeft gebracht.

(C) Deel 3

Hoe is je nieuwe niveau bepaald?

Wiski geeft een niveau aan oefeningen en leerlingen. De niveaus van zowel de leerlingen als de oefeningen veranderen door het juist of fout oplossen van oefeningen. ⓘ



Tot slot kan je nu je huidige niveau zien.

OPNIEUW ↶

NAAR DE AANBEVOLEN OEFENINGEN

(D) Deel 4

FIGUUR 4.10: Uitleg met animatie over de impact van de controle (bar-grafiek).

De resultaten van de eerste *think-aloud* studie staan in Tabel 4.2 en Tabel 4.3. De eerste toont de resultaten voor de applicatie zelf, de tweede voor de vragenlijst. Het algemene gevoel van de deelnemers na de *think-aloud* studie was positief: “Ik zou deze site wel echt gebruiken!” en “Wij krijgen geen meerkeuzevragen, maar voor het toelatingsexamens moeten we dat kunnen, dus zou het goed zijn om dat op deze manier te oefenen”.

De meeste feedback kwam over de uitleg over hoe Wiski werkt (cfr. Figuur 4.2 en Figuur 4.3). Enerzijds was de uitleg over hoe Wiski werkt te ingewikkeld en anderzijds zouden de meeste deelnemers het niet bekijken in een echt systeem. De pagina over hoe de leerling controle krijgt over de aanbevelingen was wel duidelijk voor iedereen. Daarnaast waren er ook opmerkingen over de snelheid van de animaties, zowel dat ze te snel was als te traag. Tot slot zouden verschillende deelnemers graag uitleg krijgen over waarom een oefening fout is.

Bij de vragenlijst was er slechts één probleem dat meermaals voorkwam. Verschillende deelnemers hadden niet door dat de tweede pagina over nieuwsgierigheid een vervolg was op de eerste, wat een groot probleem is aangezien de tweede pagina naar meer details vraagt over een antwoord dat gegeven werd op de eerste pagina.

TABEL 4.2: Problemen die aan het licht kwamen over de applicatie tijdens de eerste *think-aloud* studie.

Problemen	Aantal	Mogelijke oplossing
Ik snap niet alles van de uitleg over hoe Wiski werkt. Tekeningen met mondelinge uitleg zorgen ervoor dat ik het wel snap.	4/7	Tekeningen houden en uitleg herschrijven. “Hoe help ik Wiski?” was wel duidelijk, dus enkel “Hoe werkt Wiski?” aanpassen.
Ik zou niet kijken naar de uitleg over hoe Wiski werkt.	4/7	Na registratie vragen of gebruiker de uitleg wil zien. De gebruikers kunnen de werking ten alle tijden bekijken via hun profiel.
Visualisatie met één balk is onduidelijker dan de anderen. “Je kan niet goed terugkijken.”	3/7	De visualisatie heeft niet de voorkeur van de leerlingen en wordt dus niet gebruikt in de effectieve applicatie.
Ik zou graag een uitleg krijgen over waarom een oefening fout is.	3/7	Doorverwijzing naar extra uitleg toevoegen: https://hoezithet.nu/lessen
Snelheid animatie: De animaties mogen sneller na elkaar. De animatie was te snel waardoor ik het begin gemist heb en een deel van de betekenis niet begrepen heb.	3/7 2/7 1/7	Enkel animatie behouden voor grafiek. De bijhorende uitleg is statisch. Hierdoor kan je geen informatie missen door de snelheid van de animatie.
Kleurgebruik lampen: Ik zou de kleuren omgekeerd kiezen, maar deze volgorde snap ik ook. Ik zou andere kleuren nemen, focus niet op goed/slecht leggen.	3/7 2/7 1/7	Houden zoals het is.

4. ONTWIKKELING

Moeilijkheden met keuze voor “gevorderd”. Ik zou het meer naar rechts plaatsen.	2/7	Kiezen voor “Gevorderde beginner”, komt van “Advanced Beginner”.
Informatie-icoontje: Niet onmiddellijk gevonden, ik had de verwijzende tekst niet gelezen. Ik zou liever een tekst hebben met een link achter “klik hier voor meer informatie”.	2/7 1/7 1/7	Het was opvallend, dat de opmerkingen kwamen bij het eerste gebruik. Bij volgende informatie-icoontjes, werden ze onmiddellijk gevonden. Uitbreiden met een knop.
Leerling denkt dat ze volledig zelf kiest welk niveau oefeningen ze krijgt. Had niet door dat Wiski ook invloed heeft.	1/7	Zorgen dat het duidelijker wordt gemaakt bij de uitleg over hoe Wiski werkt.
Profiel: ik had het wel juist en klikte op het informatie-icoontje, maar ik zou ook nog durven klikken op het lampje zelf.	1/7	Houden zoals het is.
Zelf controle krijgen over systeem: ik denk dat controle na een reeks meer invloed heeft dan controle na een oefening.	1/7	Controle na oefening weglaten. Het was niet duidelijk dat beide soorten een ander doel hadden.
Ik vind de animaties bij de lijngrafiek en het histogram niet nodig	1/7	Hierboven is al besproken om een systeem te gebruiken waarbij de tekst statisch is.

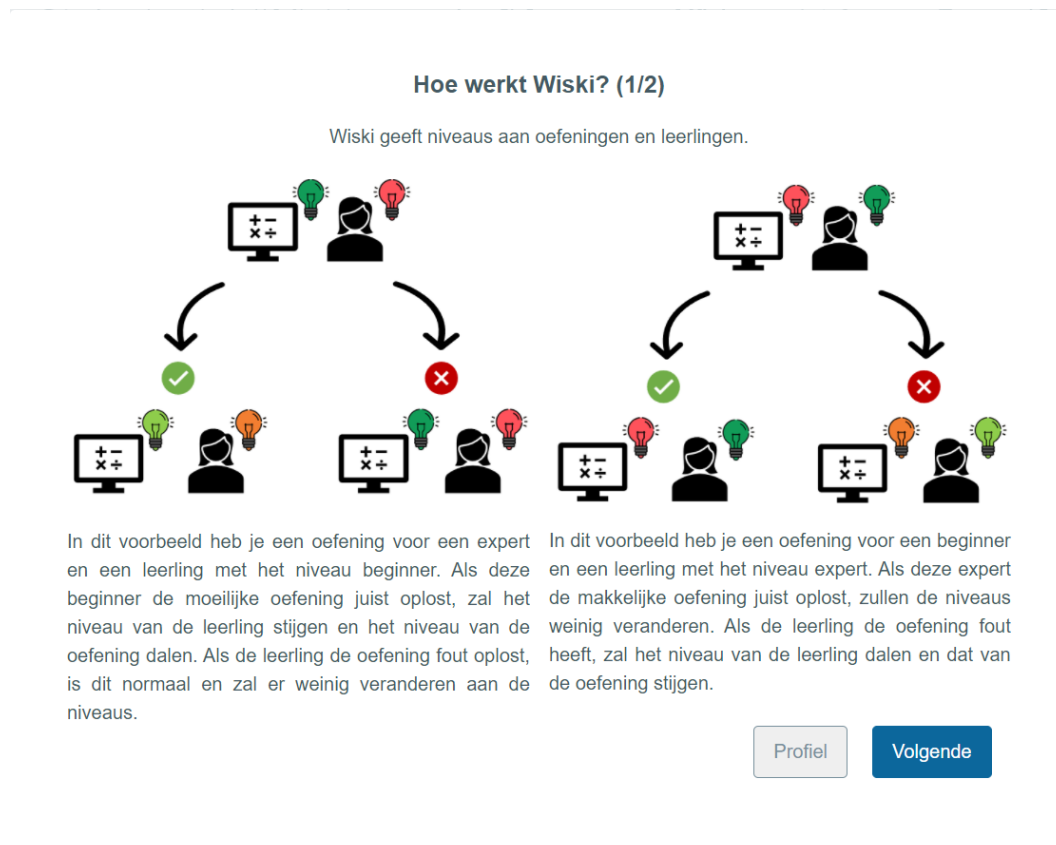
TABEL 4.3: Problemen die aan het licht kwamen over de vragenlijst tijdens de eerste *think-aloud* studie.

Problemen	Aantal	Mogelijke oplossing
Deel over nieuwsgierigheid: Ik had niet door dat het tweede deel een vervolg was op het eerste deel.	3/6	Zorgen dat het bij de uiteindelijke vragenlijst zeker op één scherm getoond wordt. Duidelijker maken dat er meer informatie wordt gevraagd i.v.m. je vorig antwoord.
“Wiski maakt gebruik van alle beschikbare kennis en informatie om aanbevelingen te produceren.” Een aantal keer moeten lezen voordat ik het begreep.	2/6	Onderscheid tussen kennis en informatie was niet duidelijk. Aanpassen naar: “Wiski maakt gebruik van de beschikbare informatie om aanbevelingen te produceren.”
Wat wordt juist bedoeld met “verklaringen”. De zin “Waarom geeft Wiski me net deze oefeningen” maakte het wel duidelijker.	1/6	Uitleg rond wat bedoeld wordt met verklaringen uitbreiden.
“Als ik verrast ben door de keuze van Wiski wil ik weten hoe dat komt.” Niet duidelijk.	1/6	“Als ik verrast ben door de keuze van Wiski wil ik weten waarom Wiski deze keuze gemaakt heeft.”
“Ik ben bekend met Wiski.” Niet duidelijk.	1/6	De zin “Ik ben vertrouwd met Wiski.” begreep de leerling wel.
“Occasioneel” begreep ik niet goed.	1/6	“Sporadisch” is een andere mogelijkheid.

4.3 *High-fidelity* prototype

Het *high-fidelity* prototype is een digitale versie van het *low-fidelity* prototype waarbij de problemen uit de eerste *think-aloud* studie zijn aangepast. Het dient daarnaast zo dicht mogelijk aan te leunen bij de finale versie, zodat zo veel mogelijk problemen opgespoord kunnen worden. Het doel van het *high-fidelity* prototype was opnieuw om vier sleutelementen te testen bij potentiële gebruikers, namelijk gebruiksvriendelijkheid, visualisatie voor uitleg over impact van controle, uitleg over werking van systeem en de vragenlijst voor het uiteindelijke onderzoek. Figuren 4.11 tot 4.18 tonen de belangrijkste elementen uit het prototype. Vier potentiële gebruikers, van 14 of 15 jaar, namen deel aan de tweede *think-aloud* studie en studeerden Wetenschappen in het derde middelbaar.

De vragen die gebruikt werden tijdens de *think-aloud* studie staan in Appendix C en de opdeling ervan is analoog aan die uit de eerste *think-aloud* studie. Ook tijdens deze studie waren er twee visualisaties en ze staan afgebeeld in Figuur 4.17 en Figuur 4.18. Er werd ook gepeild naar welke visualisatie de voorkeur kreeg: beide visualisaties werden gekozen door twee gebruikers, maar ze merkten allemaal op dat de andere optie ook goed was.



FIGUUR 4.11: Uitleg over werking Wiski (algoritme).

Hoe werkt Wiski? (2/2)

Je kan Wiski helpen met het bepalen van je niveau door je gevoel te vertellen.

Na elke reeks oefeningen kan je aangeven wat je algemeen gevoel was tijdens de reeks. Wiski zal hier dan rekening mee houden bij het bepalen van jouw nieuw niveau. Stel dat je alle vragen fout hebt opgelost, maar je denkt dat je weet waarom, dan kan je aangeven dat je toch gelijkaardige oefeningen wil. De slider hieronder zal je zien na elke reeks oefeningen.

Welk niveau van oefeningen wil je graag?

Makkelijker Gelijkaardig Moeilijker

Vorige Profiel

FIGUUR 4.12: Uitleg over werking Wiski (controle leerling).

Hoe goed denk je dat je bent in wiskunde?

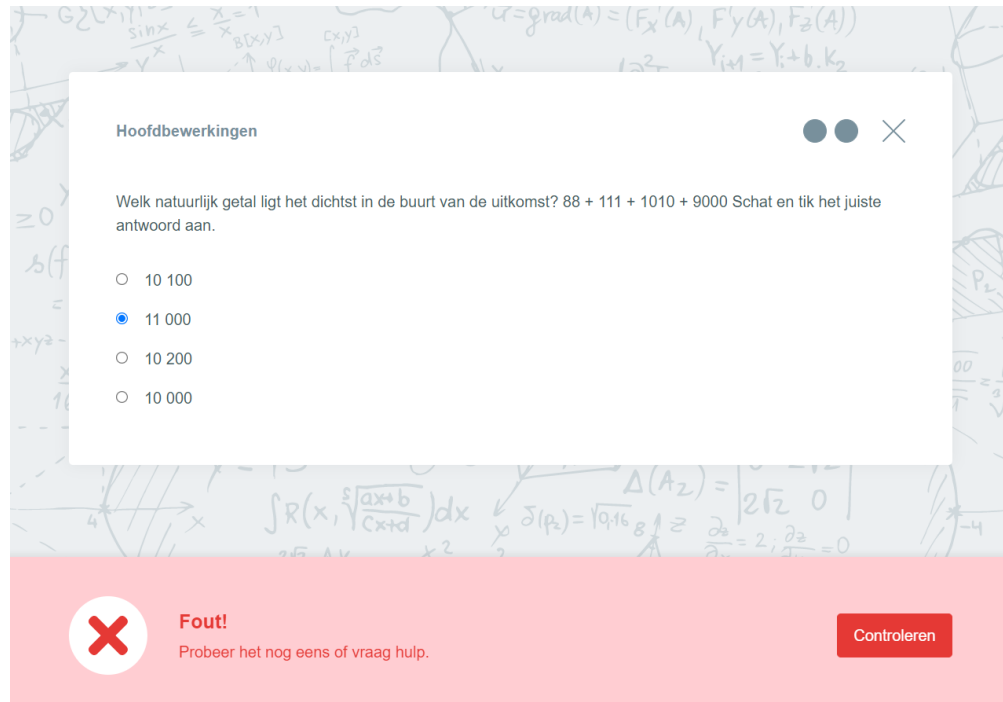
Beginner Gevorderde beginner Competent Bedreven Expert

Je geeft een inschatting van je eigen niveau voor wiskunde en er is geen juist of fout antwoord. Wiski zal je antwoord gebruiken om de juiste oefeningen voor jou te vinden.

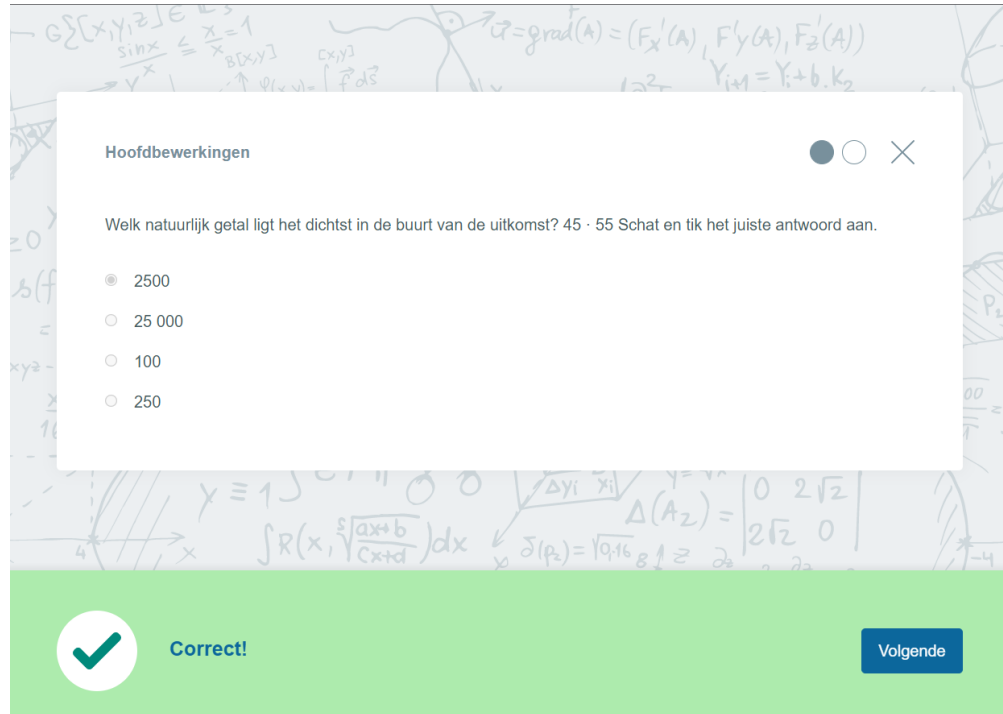
- Als beginner heb je het vaak moeilijk om wiskunde te begrijpen.
- Gevorderde beginner ben je als basisoefeningen geen probleem vormen voor jou.
- Als competente leerling scoor je gemiddeld op wiskunde.
- Als je beter scoort dan gemiddeld op wiskunde ben je bedreven.
- Voor een expert kent wiskunde geen geheimen.

Indienen

FIGUUR 4.13: Op de pagina kan de leerling een start-level kiezen.



FIGUUR 4.14: Pagina na het fout oplossen van een oefening.



FIGUUR 4.15: Pagina na het juist oplossen van een oefening.

4. ONTWIKKELING

Wiski wil graag extra informatie van jou

Je hebt een volledige reeks aanbevolen oefeningen gemaakt, proficiat! Voor de volgende reeks kan je Wiski extra informatie geven, zodat Wiski beter weet wat je gevoel is.

Welk niveau van oefeningen wil je graag?

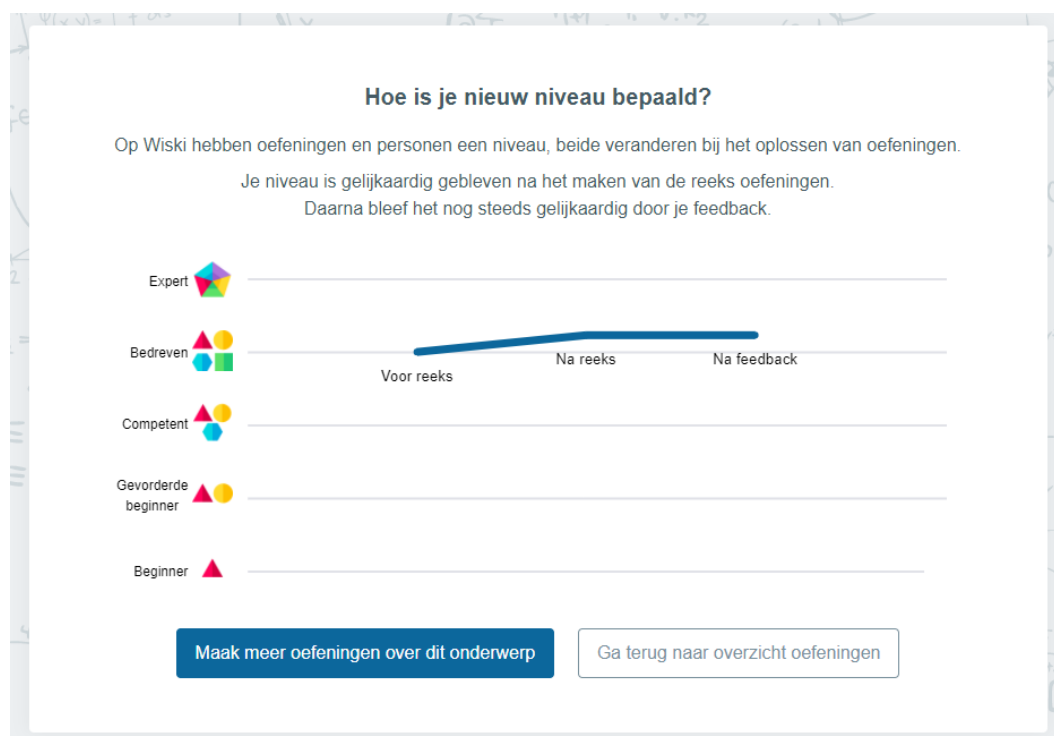
Makkelijker Gelijkwaardig Moeilijker

Indienen feedback

FIGUUR 4.16: Pagina waar de leerling kan aangeven welk niveau gewenst is voor de volgende reeks oefeningen.



FIGUUR 4.17: Uitleg over de impact van de controle (lijngrafiek).



FIGUUR 4.18: Uitleg over de impact van de controle (wiskundige symbolen).

De resultaten van de tweede *think-aloud* studie staan in Tabel 4.4 en Tabel 4.5. De eerste toont de resultaten voor de applicatie zelf, de tweede voor de vragenlijst. Het algemene gevoel van de deelnemers na de *think-aloud* studie was positief: *“Heel goede, duidelijke website. Ik zou het graag gebruiken, omdat je aan reflectie kan doen. Ik vind dit beter dan andere digitale leeromgevingen, omdat je daar je niveau niet kan aangeven.”* en *“Ik wil dit graag gebruiken. Het is makkelijk, efficiënt, heeft leuke oefeningen en het wordt ook nog eens aangepast aan mij niveau.”*

Tijdens de *think-aloud* studie waren er vier problemen die bij meerdere gebruikers opdoken. Ten eerste vonden alle vier de gebruikers het niet duidelijk dat ze opnieuw moesten proberen na een fout antwoord. Ten tweede was de uitleg over hoe het algoritme van Wiski werkt te ingewikkeld. Ten derde was de benaming van de knoppen op de pagina's over de werking van Wiski niet duidelijk. Ten slotte willen de gebruikers niet weten wie de initiële moeilijkheid van een oefening bepaald heeft, maar ze willen wel de huidige moeilijkheid van de oefening weten.

Bij de vragenlijst was er slechts één probleem dat meermaals voorkwam. Het deel over nieuwsgierigheid voor verklaringen toonde eerst een voorbeeld van een uitleg over de impact van controle dat de volledige pagina in beslag nam, daardoor wisten de gebruikers niet dat er nog vragen onder het voorbeeld stonden.

TABEL 4.4: Problemen die aan het licht kwamen over de applicatie tijdens de tweede *think-aloud* studie.

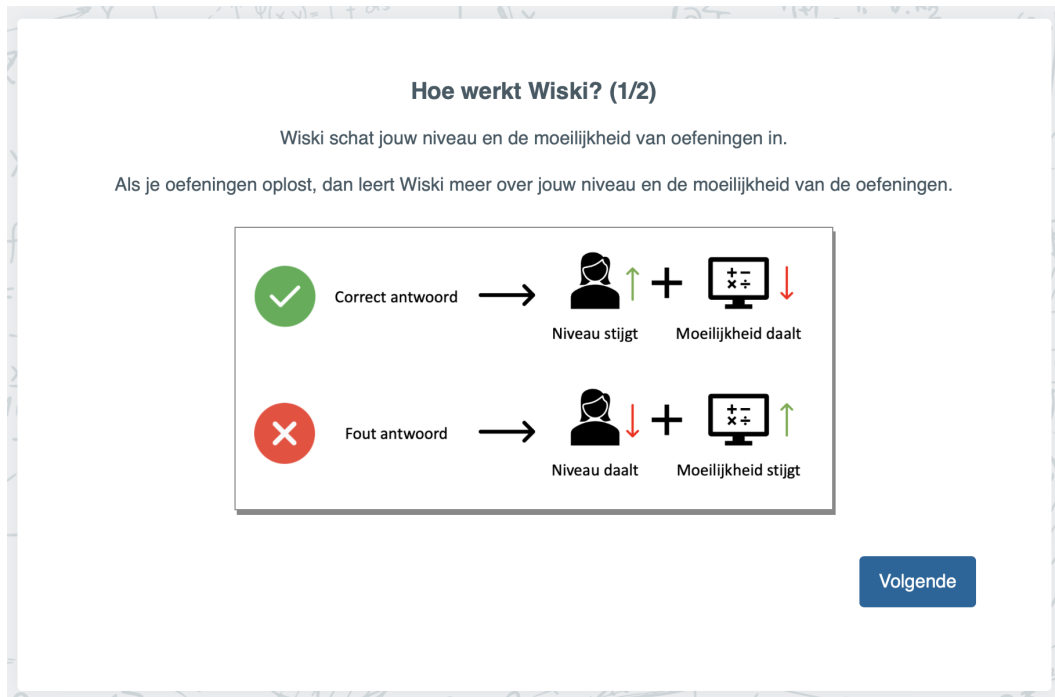
Problemen	Aantal	Mogelijke oplossing
Bij een fout antwoord is het niet duidelijk dat ik nog eens moet proberen.	4/4	Benaming knop aanpassen: “Controleren van nieuwe poging”.
De eerste pagina over de werking van Wiski is ingewikkeld.	3/4	Uitleg herschrijven: kortere tekst en afbeeldingen vereenvoudigen.
Pagina’s met werking van Wiski: de benamingen van de knoppen zijn niet duidelijk.	3/4	Andere benaming: “Uitleg afsluiten”.
Ik zou graag weten wat de moeilijkheid van de huidige oefening is.	2/4	Tijdens de oefening de moeilijkheid tonen.
Profiel: gebruiker dacht dat meer informatie over de kleur van de lamp achter de knop met meer informatie over de werking van Wiski zat.	1/4	Naam aanpassen naar “Meer informatie over werking van Wiski”.
Bij een heel laag niveau is een deel van de uitleg niet zichtbaar.	1/4	Implementatie aanpassen.

TABEL 4.5: Problemen die aan het licht kwamen over de vragenlijst tijdens de tweede *think-aloud* studie.

Problemen	Aantal	Mogelijke oplossing
Deel nieuwsgierigheid: De uitleg staat over heel het scherm, waardoor de leerlingen niet door hadden dat ze moesten scrollen om de vragen te kunnen zien.	4/4	Deel over nieuwsgierigheid maakt geen onderdeel uit van uiteindelijke vragenlijst.
Mogen kinderen onder 10 jaar de website niet gebruiken?	1/4	Aanpassen naar “jonger dan ...” en “ouder dan ...”.

4.4 Finaal *proof of concept*

Na de tweede *think-aloud* studie zijn er nog een aantal aanpassingen gebeurd aan het prototype om tot het finaal *proof of concept* te komen. Ten eerste werd gekozen voor de visualisatie met de lampjes, omdat die ook gebruikt werd tijdens de eerste *think-aloud* studie en de deelnemers toen enthousiast waren. Ten tweede werd de uitleg en de bijhorende visualisatie over het algoritme dat Wiski gebruikt vereenvoudigd, omdat het te ingewikkeld was tijdens de tweede *think-aloud* studie (cfr. Figuur 4.19). Ten derde werd de pagina waar de leerling een inschatting maakt van het eigen niveau vereenvoudigd, ook al zorgde de eerdere versie niet voor problemen bleek de vereenvoudiging toch een verbetering (cfr. Figuur 4.20). Ten slotte werd de moeilijkheid van de oefening toegevoegd aan de oefening (cfr. Figuur 4.21).



FIGUUR 4.19: Uitleg over werking Wiski (algoritme).

Hoe goed denk je dat jij bent in wiskunde?

Er is geen juist of fout antwoord. Wiski gebruikt je antwoord om de juiste oefeningen voor jou te vinden.

☐ **Expert:** wiskunde kent voor jou geen geheimen.

☐ **Bedreven:** je scoort beter dan gemiddeld op wiskunde.

☒ **Competent:** je scoort gemiddeld op wiskunde.

☐ **Gevorderde beginner:** basisoefeningen vormen geen probleem voor jou.

☐ **Beginner:** je hebt het vaak moeilijk om wiskunde te begrijpen.

Indienen

FIGUUR 4.20: Op de pagina kan de leerling een start-level kiezen.



FIGUUR 4.21: Pagina voor het maken van een oefening.

4.5 Technische implementatie

In deze paragraaf wordt de technische implementatie van Wiski besproken. Het platform is ontwikkeld met Drupal 7, een content management systeem (CMS). In een CMS wordt de inhoud opgeslagen in een gegevensbank en weergegeven via de *front-end*. Voor meer details over hoe Wiski Drupal 7 gebruikt, wordt verwezen naar het werk van Ooge [41]. De programmeertalen die verbonden zijn aan Drupal 7 zijn HTML, CSS en JavaScript voor de *front-end* en PHP voor de *back-end*. Hierna wordt dieper ingegaan op de concepten die toegevoegd zijn aan de versie van Wiski die ontwikkeld werd door Ooge.

4.5.1 Elo-classificatiesysteem

Het Elo-classificatiesysteem heeft twee hoofdcomponenten. Ten eerste het niveau van de leerling, dat wordt opgeslagen via de Userpoints module van Drupal dat een aantal nuttige functionaliteiten aanbiedt. Het heeft bijvoorbeeld een transactie systeem waardoor het niveau van de leerling doorheen de tijd bekeken kan worden. Een aanpassing aan het niveau gebeurt met een transactie. Ten tweede de moeilijkheid van elke oefening, waarvoor geen module gebruikt werd en de communicatie met de database gebeurde via SQL. Na het proberen maken van een oefening wordt de Rules module van Drupal gebruikt om het Elo-classificatiesysteem op te roepen. De nodige berekeningen om het nieuwe niveau van de leerling en de nieuwe moeilijkheid van de oefening te bepalen, gebeurt in PHP. De leerling moet blijven proberen totdat het antwoord juist is en het niveau van de leerling en de moeilijkheid van de oefening worden na elke poging aangepast.

4.5.2 Aanbevelingssysteem

Het aanbevelingssysteem bepaalt welke oefeningen het beste passen bij het niveau van de leerling en daarvoor wordt steeds een procedure doorlopen via PHP en SQL:

1. Alle potentiële oefeningen horende bij het onderwerp dat de leerling wil oefenen worden bepaald.
2. Voor elke potentiële oefening wordt de kans dat de gebruiker de oefening juist beantwoordt berekend.
3. De twee oefeningen waarvoor de kans op een juist antwoord het dichtste liggen bij 0.70 worden aanbevolen. Er is gekozen voor de waarde 0.70 omdat uit de literatuurstudie blijkt dat het enerzijds zorgt voor voldoende motivatie en anderzijds voldoende informatief is.

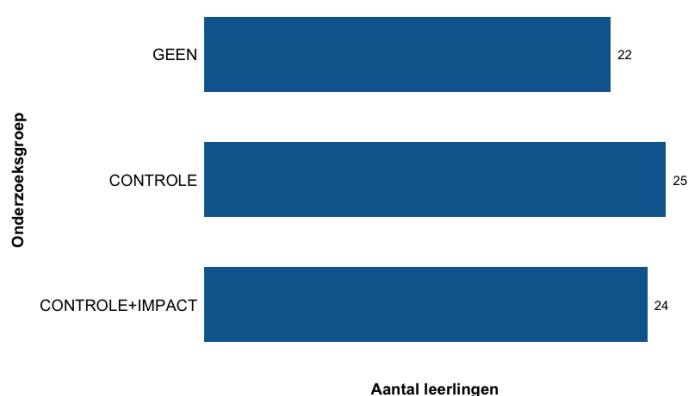
4.5.3 Verwerkingsmechanisme voor controle van gebruiker

Elke leerling krijgt na het maken van een reeks oefeningen gedeeltelijke controle over de volgende aanbevelingen. Hiervoor wordt gebruik gemaakt van een schuifbalk waarachter continue waarden zitten tussen 0.90 en 1.10. Wanneer de leerling zijn/haar controle indient, wordt PHP-code uitgevoerd die de verandering van het niveau van de leerling berekent. De verandering wordt opgeslagen via een transactie, zodat het achteraf mogelijk is om te achterhalen wat de exacte keuze van de leerling was.

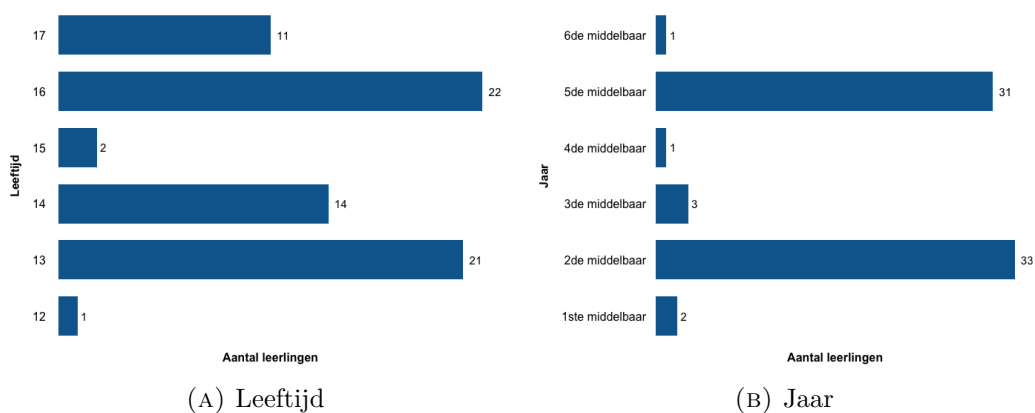
Hoofdstuk 5

Resultaten

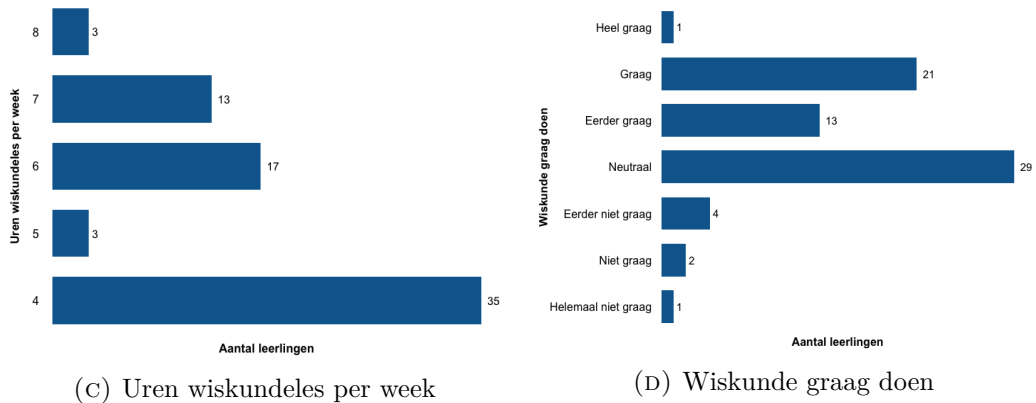
In dit hoofdstuk worden de resultaten van de gebruikersstudie besproken. In totaal namen 71 leerlingen uit het secundair onderwijs deel aan het onderzoek. Figuur 5.1 geeft de verdeling van de leerlingen over de verschillende onderzoeksgroepen weer en Figuur 5.2 toont de demografische kenmerken van de deelnemende leerlingen.



FIGUUR 5.1: Verdeling van deelnemers onderzoek in groepen.



5. RESULTATEN



FIGUUR 5.2: Demografische kenmerken van deelnemers onderzoek.

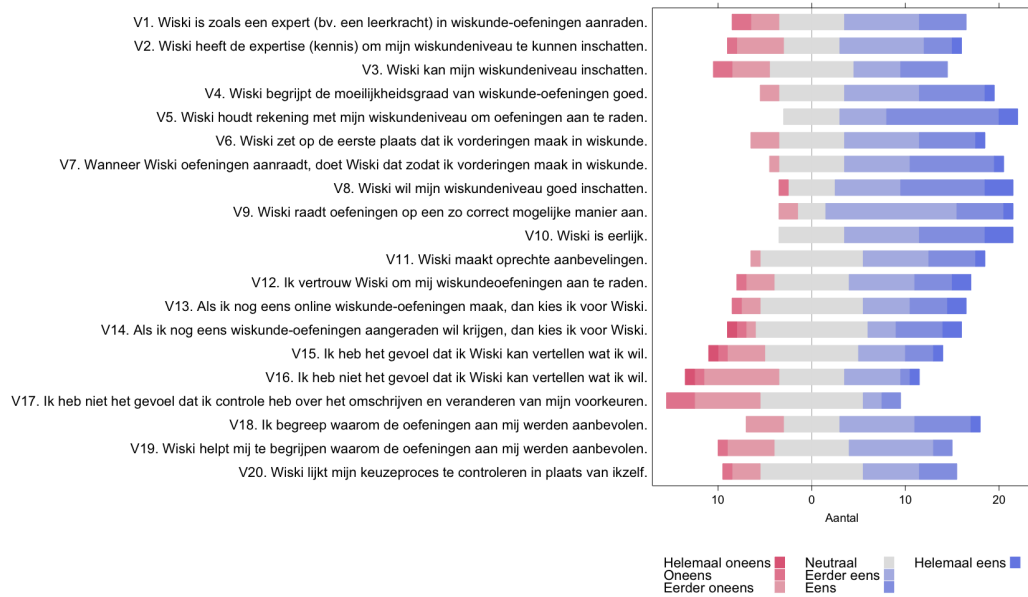
5.1 Vertrouwen in de aanbevelingen

Voor alle onderzoeksgroepen werden divergerende staafdiagrammen gemaakt over de gegeven antwoorden (cfr. Figuur 5.3). De gebruikte vragen konden verder onderverdeeld worden in verschillende aspecten van vertrouwen en de boxplots in Figuur 5.4 tonen de verdelingen van de antwoorden van de gebruikers per aspect.



(A) Antwoorden vragenlijst vertrouwen (GEEN)

5.1. Vertrouwen in de aanbevelingen



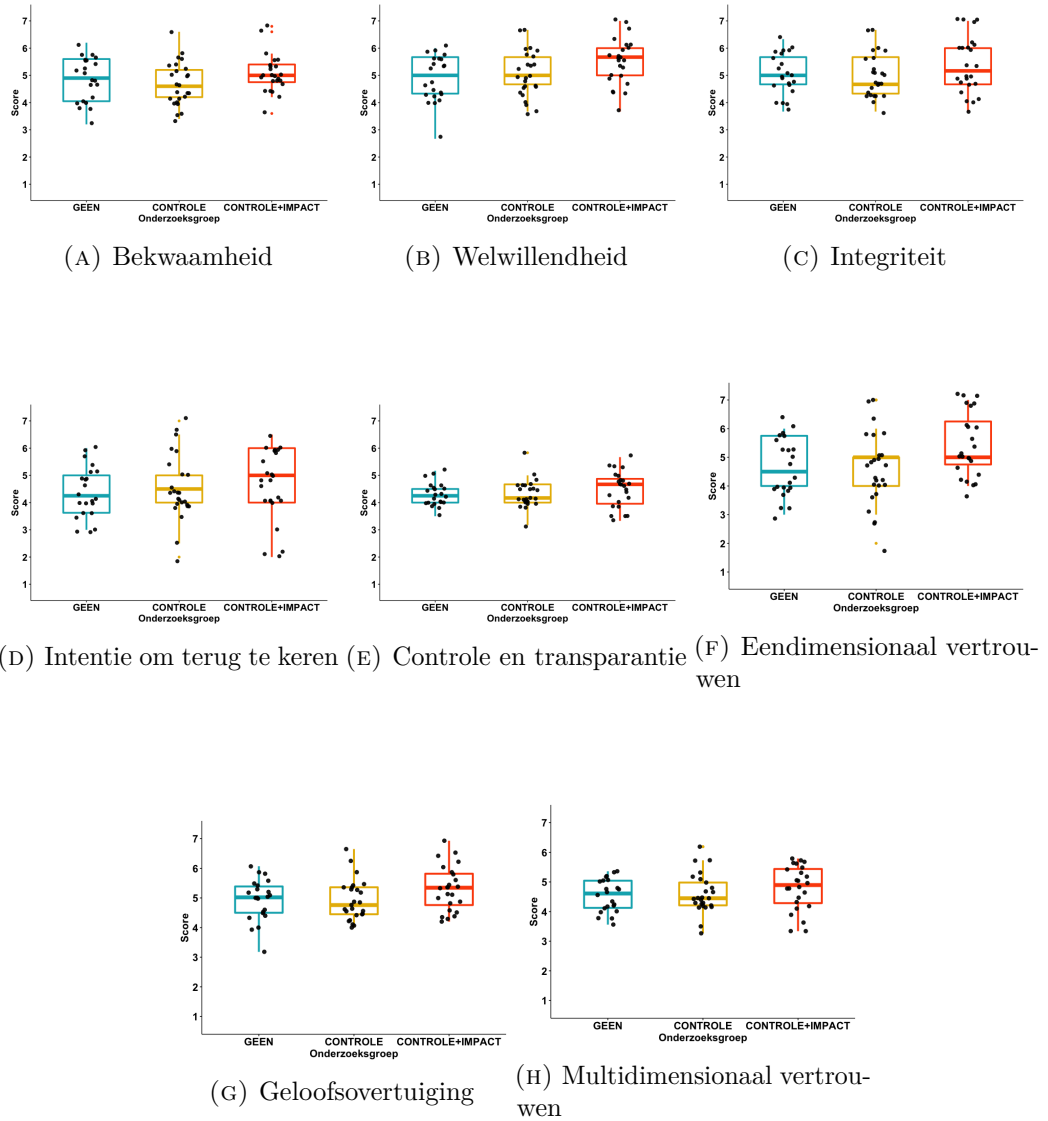
(B) Antwoorden vragenlijst vertrouwen (CONTROLE)



(C) Antwoorden vragenlijst vertrouwen (CONTROLE+IMPACT)

FIGUUR 5.3: Divergerende staafdiagrammen van antwoorden over vertrouwen in de aanbevelingen per onderzoeksgroep.

5. RESULTATEN



FIGUUR 5.4: Boxplots over de antwoorden in verband met vertrouwen in de aanbevelingen per vertrouwensaspect.

De resultaten van de analyse tussen GEEN en CONTROLE hadden allemaal een p -waarde groter dan 0,29 en waren dus voor geen enkele metriek statistisch significant ($p < 0,05$). Er kon dus niet verworpen worden dat de gemiddelden gelijk zijn.

De resultaten van de vergelijking tussen GEEN en CONTROLE+IMPACT staan in Tabel 5.1a en waren statistisch significant ($p < 0,05$) voor welwillendheid, eendimensionaal vertrouwen en geloofsovertuiging.

De resultaten van de vergelijking tussen CONTROLE en CONTROLE+IMPACT staan in Tabel 5.1b en waren statistisch significant ($p < 0,05$) voor bekwaamheid, welwillendheid, eendimensionaal vertrouwen en geloofsovertuiging.

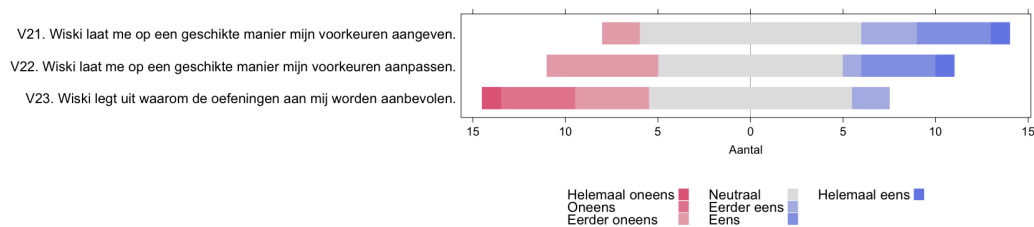
TABEL 5.1: Resultaten van statistische analyse in verband met vertrouwen in aanbevelingen per vertrouwensaspect.

(A) GEEN vs. CONTROLE+IMPACT	
	p-waarde
Bekwaamheid	0,18
Welwillendheid	0,02*
Integriteit	0,12
Intentie om terug te keren	0,16
Controle en transparantie	0,14
Eendimensionaal vertrouwen	0,02*
Geloofsovertuiging	0,04*
Multidimensionaal vertrouwen	0,10
**p < 0,01, *p < 0,05	

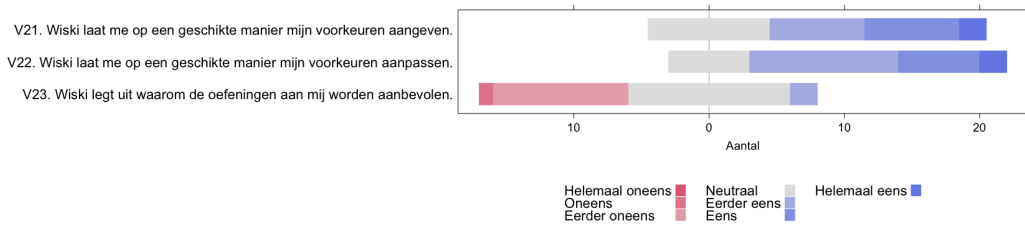
(B) CONTROLE vs. CONTROLE+IMPACT	
	p-waarde
Bekwaamheid	0,04*
Welwillendheid	0,04*
Integriteit	0,08
Intentie om terug te keren	0,33
Controle en transparantie	0,17
Eendimensionaal vertrouwen	0,02*
Geloofsovertuiging	0,03*
Multidimensionaal vertrouwen	0,16
**p < 0,01, *p < 0,05	

5.2 Interactie-adequaatheid

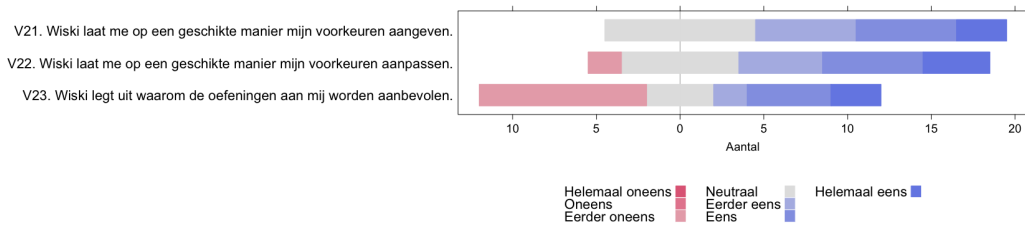
Voor alle onderzoeksgroepen werden divergerende staafdiagrammen gemaakt over de gegeven antwoorden (cfr. Figuur 5.5). De boxplots in Figuur 5.6 tonen de verdeling van de antwoorden van de gebruikers.



(A) Antwoorden vragenlijst interactie-adequaatheid (GEEN)

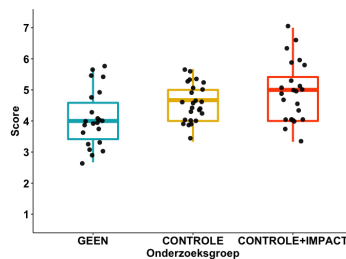


(B) Antwoorden vragenlijst interactie-adequaatheid (CONTROLE)



(C) Antwoorden vragenlijst interactie-adequaatheid (CONTROLE+IMPACT)

FIGUUR 5.5: Divergerende staafdiagrammen van antwoorden over interactie-adequaatheid per onderzoeksgroep.



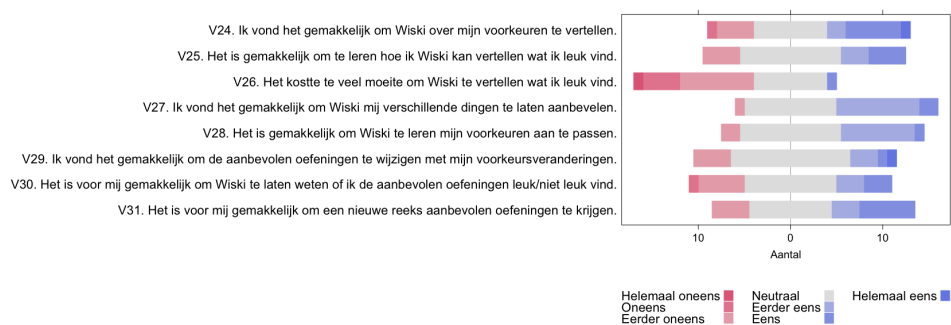
FIGUUR 5.6: Boxplots over de antwoorden in verband met interactie-adequaatheid.

De resultaten van de vergelijkingen waren statistisch significant voor GEEN vs. CONTROLE ($p = 0,01$) en GEEN vs. CONTROLE+IMPACT ($p = 0,00$).

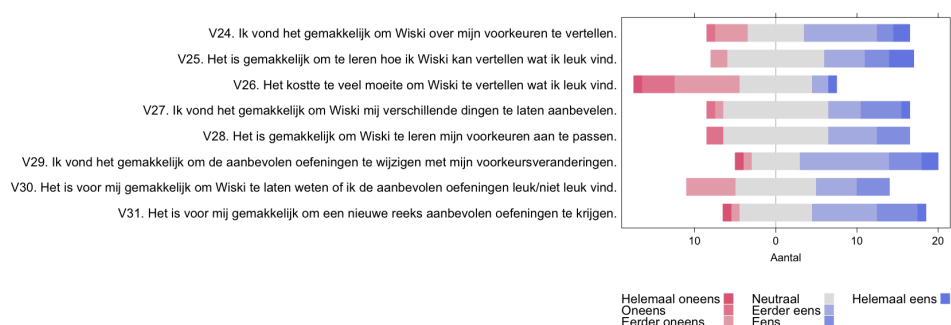
5.3 Waargenomen gebruiksgemak

Voor alle onderzoeksgroepen werden divergerende staafdiagrammen gemaakt van de gegeven antwoorden (cfr. Figuur 5.7). De gebruikte vragen konden onderverdeeld worden in verschillende aspecten van waargenomen gebruiksgemak en de boxplots in Figuur 5.8 tonen de verdelingen van de antwoorden van de gebruikers per aspect.

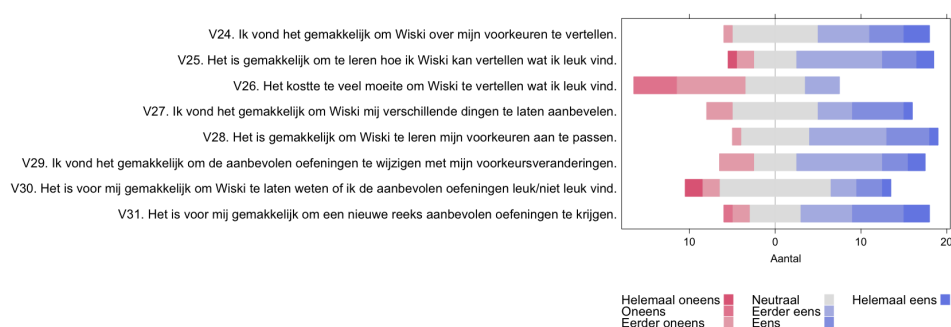
5.3. Waargenomen gebruiksgemak



(A) Antwoorden vragenlijst waargenomen gebruiksgemak (GEEN)

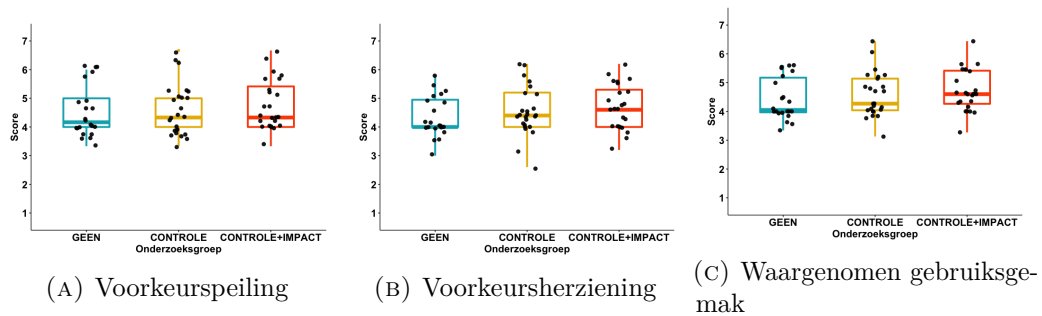


(B) Antwoorden vragenlijst waargenomen gebruiksgemak (CONTROLE)



(C) Antwoorden vragenlijst waargenomen gebruiksgemak (CONTROLE+IMPACT)

FIGUUR 5.7: Divergerende staafdiagrammen van antwoorden over waargenomen gebruiksgemak per onderzoeksgroep.



FIGUUR 5.8: Boxplots over de antwoorden in verband met waargenomen gebruiksgemak per aspect.

De resultaten van de vergelijkingen gaven voor alle metriek en alle vergelijkingen een p -waarde groter dan 0,10 en waren dus niet statistisch significant ($p < 0,05$). Er kon dus niet verworpen worden dat de gemiddelden gelijk zijn.

5.4 Kwalitatieve data

Bij elke set vragen hadden de gebruikers de mogelijkheid om extra informatie te geven bij hun antwoorden. In deze paragraaf worden de antwoorden van de gebruikers gegroepeerd per onderzoeksgroep.

5.4.1 Vertrouwen in de aanbevelingen

Na de vragen in verband met vertrouwen werd expliciet aan de gebruikers gevraagd of ze vertrouwen hebben in Wiski om oefeningen aan te bevelen, wat gebeurde aan de hand van de volgende vraag: “Heb jij vertrouwen in Wiski voor het aanbevelen van oefeningen? Waarom wel/niet?”.

GEEN. Het grootste deel (18/22) van de gebruikers die tot GEEN behoorden antwoorden positief op de vraag, wat tot uiting kwam via antwoorden zoals:

- “Ja, ik denk zeker dat als ik langer op Wiski oefen, Wiski goede oefeningen voor mij kan aanbevelen.”
- “Ja, de website ziet er professioneel en betrouwbaar uit. Ik heb ook het gevoel dat de oefeningen een goed niveau hebben!”
- “Ja, Ik denk dat Wiski goed op jouw niveau oefeningen kan aanbevelen.”

Er waren drie gebruikers die deels vertrouwen hadden in Wiski voor het aanbevelen van oefeningen, dat blijkt uit de volgende antwoorden:

- “Ik heb te weinig kunnen oefenen en de site kunnen uitvoeren, dus kan ik ook geen goed eindoordeel geven (op dit moment). Van hetgeen dat ik heb gezien, heeft deze site wel potentieel om een goede manier van oefenen voor wiskunde te worden.”

- *“Als je er lang genoeg op oefent misschien wel.”*
- *“Een beetje omdat ik vaak dezelfde oefeningen krijg.”*

Eén gebruiker gaf aan geen vertrouwen te hebben in Wiski voor het aanbevelen van oefeningen. Het antwoord van de gebruiker was: *“Nee, ze zijn iets te makkelijk ook al ben ik geen expert.”*

CONTROLE. Ook hier gaf het grootste deel (15/25) van de gebruikers aan vertrouwen te hebben in Wiski voor het aanbevelen van oefeningen. Een aantal bijhorende antwoorden zijn:

- *“Ja, het lijkt me betrouwbaar op het eerste zicht en het stelt ook goede vragen die aangepast zijn op jouw wiskundig niveau.”*
- *“Ja, omdat Wiski weet wanneer ik wat moeilijkheden heb bij oefeningen en wanneer niet. Hierdoor worden moeilijke oefeningen nog eens aanbevolen.”*
- *“Ja, omdat als je oefent, dan weet Wiski in welk niveau je zit en als je op dat niveau blijft oefenen gaat je niveau stijgen en zo leer je op een korte tijd veel dingen.”*

Acht gebruikers gaven aan deels vertrouwen te hebben in Wiski om oefeningen aan te bevelen, wat tot uiting kwam via antwoorden zoals:

- *“Wiski houdt goed rekening met wat je juist of fout doet, maar tegelijkertijd zou het ook gewoon een AI zijn om het uit te rekenen. Wiski is goed, maar ik vraag liever advies aan een fysiek persoon. Ik denk ongeveer neutraal.”*
- *“Het is een programma en geen leerkracht. Ik vertrouw het dus niet helemaal, maar voor gewoon eventjes wat oefeningen te maken is het heel handig.”*
- *“Ik denk dat het wel een betrouwbare bron is om te oefenen, maar soms hebben leerkrachten hun manier van vragen stellen en dit kan niet altijd overeenkomen met de oefeningen die Wiski aanbiedt. Het is wel een goede manier om je basis onder de knie te krijgen.”*

Eén gebruiker gaf aan geen vertrouwen te hebben in Wiski voor het aanbevelen van oefeningen: *“Wiski houdt denk ik geen rekening met autisme.”* Het antwoord van één gebruiker ontbreekt, omdat het geen antwoord was op de gestelde vraag: *“Het zijn soms rare oefeningen.”*

CONTROLE+IMPACT. Ook CONTROLE+IMPACT gaf voornamelijk aan vertrouwen te hebben in Wiski voor het aanbevelen van oefeningen (17/24), wat blijkt uit de volgende antwoorden:

- *“Ja, ik denk dat Wiski een goede manier is om oefeningen te maken. Het is heel handig dat je door juiste antwoorden moeilijkere vragen krijgt en door foute antwoorden ook terug dezelfde vragen. Je kan zo opklimmen naar een hoger niveau van oefeningen en de leerstof dus beter beheersen.”*

- *“Ja, ik denk dat Wiski wel oefeningen op mijn niveau geeft. Het is leuk dat wanneer je veel oefeningen juist hebt dat je dan moeilijkere oefeningen krijgt om jezelf uit te dagen. Je merkt dat ze moeilijker worden, hierdoor vertrouw ik de oefeningen die worden aanbevolen worden per niveau.”*
- *“Ja, handig dat dit platform je niveau kan inschatten, de oefeningen die Wiski aanbeveelt zijn daarom goed ingeschat.”*

Voor deze groep gaven drie gebruikers aan deels vertrouwen te hebben in Wiski om oefeningen aan te bevelen. Een aantal bijhorende antwoorden zijn:

- *“Ja en nee, de ene oefening vind ik makkelijk en de andere is dan weer wat moeilijker ook al zou ik die oefening perfect moeten kunnen maken.”*
- *“Hangt er van af, in de meeste gevallen passen de oefeningen wel bij het niveau. Maar soms komen er 2 of 3 keer dezelfde oefeningen voor.”*

Tot slot gaven twee gebruikers aan geen vertrouwen te hebben in Wiski voor het aanbevelen van oefeningen. Er was één gebruiker die geen antwoord gegeven heeft. De antwoorden van de gebruikers die geen vertrouwen hadden, waren:

- *“Niet echt, ik heb nu enkele oefeningen gemaakt en heb nog niet het niveau gevonden dat bij mij past. Dus ik ben meer geneigd om oefeningen in mijn boek te maken omdat ik weet dat we dat niveau moeten kunnen.”*
- *“Ik vind het moeilijk om hier een mening over te vormen, omdat ik het net één keer geprobeerd heb. Ik denk dat Wiski wel goed gemaakt is en zijn best doet om je te helpen maar ik denk niet dat het echt mijn wiskunde niveau kan bepalen.”*
- *“Wel want de dingen die ze me aanraden vind ik gemakkelijk voor dat niveau.”*

5.4.2 Interactie-adequaatheid en waargenomen gebruiksgemak

Na de vragen over interactie-adequaatheid en waargenomen gebruiksgemak hadden de gebruikers de optie om extra opmerkingen te geven. De antwoorden op de vraag worden per onderzoeksgroep besproken.

GEEN. Er waren tien gebruikers die expliciet aangaven geen opmerkingen te hebben. Twee gebruikers hadden wel een opmerking en dezelfde: *“Een aantal van de elementen die bevraagd zijn, zoals mijn voorkeur aanpassen, heb ik niet gedaan.”*

CONTROLE. Zeven gebruikers gaven expliciet aan geen opmerkingen te hebben. Twee gebruikers uit CONTROLE hadden wel een opmerking:

- *“Beetje te weinig ervaring met Wiski om dit direct in te vullen.”*
- *“Ik heb in het begin bij die balk dat je de moeilijkheidsgraad kon aanduiden na elke oefening de moeilijkheidsgraad op makkelijk gezet omdat ik dacht dat je moest aanduiden hoe je de oefening vond. Het was dus anders.”*

CONTROLE+IMPACT. Voor CONTROLE+IMPACT gaven negen gebruikers expliciet aan geen opmerkingen te hebben. Eén gebruiker had een opmerking: *“Veel dezelfde vragen op een andere manier gesteld.”*

5.4.3 Controle krijgen van Wiski

Na de vragenlijsten was er nog een optioneel deel waar de gebruikers hun mening konden geven over gedeeltelijke controle krijgen over de aanbevelingen.

GEEN. GEEN kreeg geen controle tijdens het gebruik van Wiski. Twaalf gebruikers zagen de toevoeging van de functionaliteit als een meerwaarde, wat tot uiting kwam via antwoorden zoals:

- *“Ja, omdat ik wil bevestigen dat mijn oefeningen goed zijn.”*
- *“Ja. Dit biedt de mogelijkheid om zelf ook een beetje richting te geven aan welke oefeningen die je wilt. Ook als je het wat minder doet, dat je toch ook eens moeilijkere oefeningen krijgt om te zien wat die inhouden.”*
- *“Ja, dan kan je laten weten wat je er van vindt.”*

Eén gebruiker gaf aan dat het niet uit maakt of er controle is voor de gebruiker: *“Voor mij maakt dat niet zoveel uit. Misschien is het wel goed om feedback te kunnen geven.”*

CONTROLE. CONTROLE kreeg controle tijdens het gebruik van Wiski. Elf gebruikers gaven aan dat de functionaliteit een meerwaarde was. Een aantal bijhorende antwoorden zijn:

- *“Ik vind het goed dat je dit kunt aanpassen als je aanvoelt dat je de oefeningen te moeilijk of te makkelijk vindt.”*
- *“Ik vind het een zeer handige functie van Wiski.”*
- *“Je kon na iedere 2 oefeningen zeggen wat voor moeilijkheidsgraad je wou en dat is heel handig als je eens iets moeilijker wilt proberen maar ook eens voor iets makkelijk te gaan.”*

Eén gebruiker had de schuifbalk verkeerd ingevuld: *“Ik had dit verkeerd ingevuld.”*

CONTROLE+IMPACT. CONTROLE+IMPACT kreeg gedeeltelijke controle over de aanbevelingen tijdens het gebruik van Wiski. Tien gebruikers gaven aan dat de functionaliteit een meerwaarde was. Het kwam tot uiting met antwoorden zoals:

- *“Ik vind het een heel goede toepassing dat je zelf ook kan zeggen welke niveau van oefeningen je wilt. Als je dan zelf nog een beetje wil oefenen op basis oefeningen kan dat ook nog steeds terwijl dat wel onder je niveau zou zijn.”*
- *“Ik vond de methode wel handig waarmee je je niveau kon aanpassen. Dit was allemaal duidelijk aangegeven, het was gewoon moeilijk om het perfecte niveau te vinden.”*

- *“Ik vind het goed dat je na een oefening kan kiezen of je gelijkaardige of andere oefeningen wilt. Zo blijf je niet vastzitten in te moeilijke oefeningen. De 3 keuzes zijn voor mij voldoende.”*

Eén gebruiker gaf de volgende opmerking: *“Ik zei dat ik makkelijkere oefeningen wilde maar ik kreeg exact dezelfde dus dat klopte niet.”*

5.4.4 Uitleg over impact van controle

Na de vragenlijsten was er nog een optioneel deel waar de gebruikers hun mening konden geven over de uitleg van Wiski over de impact van de controle.

GEEN. GEEN kreeg geen uitleg tijdens het gebruik van Wiski. De groep kreeg ook geen controle, waarop de uitleg gebaseerd wordt. Veertien gebruikers zouden het een meerwaarde vinden. Een aantal bijhorende antwoorden zijn:

- *“Dit kan op verschillende manieren handig zijn om te zien waarom je op een bepaald niveau bent. Ik zou zeker dit of iets gelijkaardig invoeren.”*
- *“Dit zou wel handig zijn, dan kan je zien hoe je evolueert.”*
- *“Ja, dat geeft wat meer overzicht of je geëvolueerd bent.”*

CONTROLE. CONTROLE kreeg geen uitleg tijdens het gebruik van Wiski. De groep kreeg wel controle, waarop de uitleg gebaseerd zou zijn. Achttien gebruikers vonden de toevoeging van een uitleg over de impact een meerwaarde. Een aantal bijhorende antwoorden zijn:

- *“Ja, dat zou handig zijn omdat je dan weet hoe je vordert op Wiski.”*
- *“Ja, want dan kan je weten of je vooruitgang maakt en/of je achteruitgang maakt.”*
- *“Ik denk dat dit een heel goede manier is maar ik denk dat het ook het best is dat Wiski het ook weergeeft in cijfers.”*

Eén gebruiker antwoordde nee op de vraag, zonder extra uitleg.

CONTROLE+IMPACT. CONTROLE+IMPACT kreeg wel uitleg over de impact van de controle tijdens het gebruik van Wiski. Vijf gebruikers gaven aan het een meerwaarde te vinden.

- *“De grafiek is heel duidelijk, je krijgt een goed overzicht over je niveau.”*
- *“De verklaringen zijn handig om je evolutie te zien, het is ook duidelijk maar simpel weergegeven.”*
- *“Heeft een mooie lay-out en is leuk gedaan.”*

Twee gebruikers hadden een opmerking over de huidige wijze: de ene wou graag meer uitleg (*“Iets meer uitleg hierover.”*), de andere vond de kloof tussen expert en bedreven te groot (*“Iets te grote kloof tussen expert en bedreven”*).

Hoofdstuk 6

Discussie

In dit hoofdstuk worden de resultaten uit Hoofdstuk 5 geïnterpreteerd en gekoppeld aan bevindingen in de literatuur. Op basis van de discussie worden de onderzoeksvragen uit Hoofdstuk 3 beantwoord. Waargenomen gebruiksgemak wordt als eerste besproken, omdat het een belangrijke invloed heeft op de resultaten van interactie-adequaatheid en vertrouwen. Daarna worden interactie-adequaatheid en vertrouwen in de aanbevelingen besproken. Vervolgens wordt toegelicht hoe leerlingen en leerkrachten een digitale leeromgeving, zoals Wiski, graag zouden gebruiken. Tot slot worden de beperkingen van het onderzoek besproken.

6.1 Waargenomen gebruiksgemak

Waargenomen gebruiksgemak meet de mogelijkheid van de gebruiker om snel en correct taken uit te voeren. Voor geen enkele vergelijking van groepen is er statistisch significant bewijs dat er een verschil is tussen de groepen. Er kan dus besloten worden dat het toevoegen van de functionaliteiten er niet voor gezorgd heeft dat het moeilijker is om de website te gebruiken. Het is een belangrijk besluit omdat het betekent dat de resultaten in verband met interactie-adequaatheid en vertrouwen niet beïnvloed zijn door het waargenomen gebruiksgemak.

Het onderzoek van Mabbott et al. [33] toonde aan dat leerlingen zich minder comfortabel voelen wanneer ze volledige controle hebben over het model dan wanneer ze suggesties mogen doen, maar het systeem de eindbeslissing heeft. De gebruikers van Wiski kregen gedeeltelijke controle en de resultaten tonen niet aan dat ze zich hier oncomfortabel bij voelen, wat aansluit bij het resultaat van Mabbott et al.

6.2 Interactie-adequaatheid

Interactie-adequaatheid omvat de integratie van drie mechanismen: voorkeursuitlokking, voorkeursherziening en verklaring van resultaten. GEEN kon enkel gebruikmaken van het mechanisme voor voorkeursuitlokking, doordat de leerlingen hun eigen startniveau konden kiezen tijdens de registratie. CONTROLE kon de mechanismen voor voorkeursuitlokking en voorkeursherziening gebruiken, door het eigen startniveau te

kiezen en de controle die ze krijgen over de aanbevelingen. CONTROLE+IMPACT kon de eerste twee zeker gebruiken, over het derde bestaat twijfel. Het interactiemechanisme ‘verklaring van de resultaten’ werd bevraagd door “Wiski legt uit waarom de oefeningen aan mij worden aanbevolen”. Voor de toevoeging van de uitleg over de impact in vergelijking met enkel controle is er geen statistisch significant bewijs dat aantoonst dat interactie-adequaatheid verhoogt. De gebruikers zien de uitleg dus niet als een verklaring voor de aanbevelingen. De resultaten van de vergelijkingen tussen GEEN en CONTROLE enerzijds en GEEN en CONTROLE+IMPACT anderzijds tonen wel statistisch significant bewijs dat de toevoeging van controle of de toevoeging van controle en een uitleg over de impact ervan interactie-adequaatheid verhogen.

Uit de resultaten blijkt dus dat de uitleg van Wiski over de impact van de controle als doel heeft om te verklaren wat er gebeurd is met de controle van de gebruiker. Het doel van XAI is meestal anders, namelijk een verklaring bieden over de besluiten en acties van het systeem met het oog op meer transparantie en vertrouwen [3, 6]. Vaak omvat het een verklaring over waarom iets aanbevolen wordt.

Volgens Ooge et al. [42] kunnen de technieken voor verklaringen opgedeeld worden in drie categorieën, op basis van hoe ze het black-box-probleem openen: het model zelf uitleggen, de uitkomsten uitleggen of het model inspecteren. De resultaten tonen dat Wiski het black-box-probleem niet opent door uitkomsten uit te leggen, maar het probeert de andere twee wel. Door middel van de uitleg over de impact van de controle geeft Wiski de leerlingen de mogelijkheid om hun model te bekijken en te begrijpen, wat onderbouwd wordt door volgende citaten van leerlingen: *“Ja, want dan kan je weten of je vooruitgang maakt en/of je achteruitgang maakt.”*, *“Dit kan op verschillende manieren handig zijn om te zien waarom je op een bepaald niveau bent.”* en *“De verklaringen zijn handig om je evolutie te zien, het is ook duidelijk maar simpel weergegeven.”*

6.3 Vertrouwen in de aanbevelingen

Hierna worden de resultaten in verband met vertrouwen in de aanbevelingen, gegroepeerd per functionaliteit, besproken. Daarna worden algemene connecties met de literatuur toegelicht.

6.3.1 Controle over aanbevelingen

De resultaten in verband met vertrouwen tonen geen statistisch significant bewijs dat aantoonst dat het toevoegen van de controle het vertrouwen van de gebruiker verhoogt. De gebruikers moesten op het einde een open vraag beantwoorden in verband met vertrouwen, namelijk “Heb jij vertrouwen in Wiski voor het aanbevelen van oefeningen? Waarom wel/niet?”. Er waren twee thema’s die terugkwamen bij de antwoorden. Ten eerste werd meermaals aangegeven dat de vragen op Wiski niet aangepast zijn aan de vraagstelling van de eigen leerkracht, die leerlingen geven de voorkeur aan pen-en-papier oefeningen om een toets voor te bereiden. Ten tweede gaven verschillende leerlingen aan dat ze het moeilijk vonden om een mening te vormen na een beperkt gebruik van Wiski. De leerlingen konden ook hun mening

geven over de controle over de aanbevelingen. Bij GEEN gaf het grootste deel aan dat ze de toevoeging een meerwaarde zouden vinden. Bij CONTROLE gaf het grootste deel aan dat het een handige functionaliteit is. Toch heeft de toevoeging ervan, na een korte studieperiode, geen invloed op het vertrouwen in de aanbevelingen.

6.3.2 Uitleg over impact van controle

De resultaten tonen dat het toevoegen van de controle en de uitleg over de impact ervan een statistisch significant effect hebben op het verhogen van de vertrouwensaspecten welwillendheid, eendimensionaal vertrouwen en geloofsovertuiging, in vergelijking met geen functionaliteiten. Daarnaast tonen de resultaten ook dat de toevoeging van controle en de uitleg over de impact ervan een statistisch significant effect hebben op het verhogen van de vertrouwensaspecten bekwaamheid, welwillendheid, eendimensionaal vertrouwen en geloofsovertuiging, in vergelijking met enkel controle. Als welwillendheid hoger is, geloven de gebruikers dat Wiski om hen geeft en in hun belang handelt. Als bekwaamheid hoger is, geloven de gebruikers dat Wiski het potentieel, de kunde en de expertise heeft om goede aanbevelingen te leveren.

Er zijn echter geen statistisch significante aanwijzingen dat multidimensionaal vertrouwen verhoogt door de toevoeging van de functionaliteiten. Volgens de literatuur is vertrouwen moeilijk te meten/definiëren [43] en kan één expliciete vraag de complexiteit ervan moeilijk weergeven. Het is mogelijk dat bepaalde leerlingen andere elementen van het systeem in overweging namen tijdens het beantwoorden van de vraag voor eendimensionaal vertrouwen, terwijl multidimensionaal vertrouwen verschillende aspecten bevraagt. Het is dus moeilijk om te besluiten of het vertrouwen in de aanbevelingen hoger is, enkel op basis van eendimensionaal vertrouwen. Het feit dat geloofsovertuiging hoger is, geeft wel een beter beeld aangezien het samengesteld is uit bekwaamheid, welwillendheid en integriteit en dus meerdere aspecten van het systeem bevraagd heeft. Toch zijn niet alle bevraagde aspecten van vertrouwen hoger, waardoor een genuanceerd antwoord op de vraag of vertrouwen in aanbevelingen verhoogt door het toevoegen van de functionaliteiten nodig is.

6.3.3 Algemene connecties met literatuur

De bovenstaande bevindingen hebben een aantal connecties met bestaande literatuur.

Ten eerste worden metacognitieve kennis [40] en zelfregulerende leerprocessen (metacognitief proces en gedragsproces) [53, 60] bevordert door het gebruik van Wiski. De controle en de bijhorende uitleg over de impact ervan zorgen ervoor dat leerlingen bewuster nadenken over welk niveau van oefeningen de voorkeur krijgt en hoe hun niveau verandert, wat zowel zelfreflectie als reflectie van het leerproces bevordert. De bevordering van zelfreflectie wordt onderbouwd door *“Ik vind het goed dat je dit kunt aanpassen als je aanvoelt dat je de oefeningen te moeilijk of te makkelijk vindt.”* en *“Ik vind het een heel goede toepassing dat je zelf ook kan zeggen welke niveau van oefeningen je wilt. Als je dan zelf nog een beetje wil oefenen op basis oefeningen kan dat ook nog steeds terwijl dat wel onder je niveau zou zijn.”*. De bevordering van reflectie van het eigen leerproces wordt onderbouwd door *“Ik denk dat het wel*

een betrouwbare bron is om te oefenen, maar soms hebben leerkrachten hun manier van vragen stellen en dit kan niet altijd overeenkomen met de oefeningen die Wiski aanbiedt. Het is wel een goede manier om je basis onder de knie te krijgen.” en “Niet echt, ik heb nu enkele oefeningen gemaakt en heb nog niet het niveau gevonden dat bij mij past. Dus ik ben meer geneigd om oefeningen in mijn boek te maken omdat ik weet dat we dat niveau moeten kunnen.”.

Ten tweede hebben OLMs een aantal positieve effecten [7]: (1) accuraatheid van het model verhoogt, (2) zelf-reflectie wordt gepromoot, (3) leerlingen krijgen meer controle over hun leerproces en (4) het vertrouwen in het systeem verhoogt. De resultaten tonen dat al de effecten minstens gedeeltelijk bevestigd worden door de gebruikers. De accuraatheid van het model wordt verhoogd doordat de gebruikers hun gevoel kunnen overbrengen aan Wiski. Er wordt niet enkel rekening gehouden met het resultaat van de oefeningen, maar ook met de wensen van de gebruiker, waardoor het model past bij zowel de objectieve noden als de subjectieve wensen van de gebruiker. Het volgende citaat ondersteunt de bevinding: *“Ik vind het een heel goede toepassing dat je zelf ook kan zeggen welke niveau van oefeningen je wilt. Als je dan zelf nog een beetje wil oefenen op basis oefeningen kan dat ook nog steeds terwijl dat wel onder je niveau zou zijn.”* Zelf-reflectie wordt gepromoot doordat de gebruikers nadenken over hun wensen en hun evolutie kunnen volgen door de uitleg over de impact van de controle (cfr. *“De verklaringen zijn handig om je evolutie te zien, het is ook duidelijk maar simpel weergegeven.”*). De gebruikers krijgen meer controle over hun leerproces en gaven dat tijdens het onderzoek ook zelf aan: *“Ik vind het goed dat je na een oefening kan kiezen of je gelijkaardige of andere oefeningen wilt. Zo blijf je niet vastzitten in te moeilijke oefeningen. De 3 keuzes zijn voor mij voldoende.”* Uit de resultaten blijkt dat het vertrouwen van de gebruikers verhoogt voor een aantal aspecten van vertrouwen.

Ten derde hebben de leerlingen van dit onderzoek niet aangegeven dat het openstellen van hun model de duidelijkheid van het systeem verlaagt, een gevoel van oneerlijkheid oproept of zijn/haar engagement verlaagt, wat wel nadelen waren tijdens het onderzoek van Abdi et al. [2].

Ten vierde werd bij de discussie over interactie-adequaatheid het verschil besproken tussen de uitleg over de impact van de controle en XAI. De resultaten in verband met vertrouwen versterken dat verschil doordat er geen statistisch significant verschil was op vlak van controle en transparantie. Er is wel een statistisch significant verschil op vlak van bepaalde vertrouwensaspecten.

Ten slotte waren ook bij het onderzoek van Kato [26] de resultaten in verband met eendimensionaal vertrouwen anders dan voor multidimensionaal vertrouwen. Alle onderzoeksgroepen scoorden eendimensionaal vertrouwen hoog, waardoor er geen statistisch significant verschil was tussen de groepen. Hij wijt dat aan de invloed van andere elementen van het platform, zoals bijvoorbeeld de kwaliteit van de oefeningen. Op het vlak van multidimensionaal vertrouwen was er wel een statistisch significant verschil. De voorgaande bevindingen ondersteunen de bespreking in verband met eendimensionaal en multidimensionaal vertrouwen uit deze discussie: om een genuanceerd antwoord te kunnen geven in verband met het complexe begrip vertrouwen moet het onderzocht worden op basis van verschillende aspecten.

6.4 Digitale leeromgevingen in het onderwijs

In deze paragraaf wordt uitgezoomd naar de bredere context van digitale leeromgevingen in het onderwijs. Uit de pilootstudie en de gebruikersstudie kwamen enerzijds de context waarin het nuttig is en anderzijds de belangrijkste eigenschappen voor een digitale leeromgeving en voor het bijhorende aanbevelingssysteem naar voren.

Zowel de leerkrachten als de leerlingen geven aan dat ze een digitale leeromgeving niet willen gebruiken in verband met evaluatie. De leerkrachten vinden meerkeuzevragen het probleem, aangezien de werkwijze van de leerlingen niet gevraagd wordt. Meerkeuzevragen vormen voor de leerlingen geen probleem, één leerling gaf tijdens een *think-aloud* studie aan dat het een goede voorbereiding is op het toelatingsexamen (cfr. *“Wij krijgen geen meerkeuzevragen, maar voor het toelatingsexamens moeten we dat kunnen, dus zou het goed zijn om dat op deze manier te oefenen”*). De leerlingen willen een dergelijk platform niet gebruiken om evaluatie voor te bereiden, omdat de vragen bijvoorbeeld niet aangepast zijn aan de vraagstelling van de leerkrachten, wat blijkt uit citaten zoals *“Ik denk dat het wel een betrouwbare bron is om te oefenen, maar soms hebben leerkrachten hun manier van vragen stellen en dit kan niet altijd overeenkomen met de oefeningen die Wiski aanbiedt. Het is wel een goede manier om je basis onder de knie te krijgen.”* en *“Het is een programma en geen leerkracht. Ik vertrouw het dus niet helemaal, maar voor gewoon eventjes wat oefeningen te maken is het heel handig.”* Tijdens de pilootstudie gaven de leerkrachten aan dat ze graag zelf oefeningen zouden toevoegen aan het platform. Indien dat mogelijk zou zijn, zou het kunnen dat leerlingen een dergelijk platform wel willen gebruiken om evaluatie voor te bereiden. Beide groepen willen een digitale leeromgeving, zoals Wiski, dus gebruiken voor basiskennis en remediëring. De context waarin zowel leerkrachten als leerlingen een digitale leeromgeving willen gebruiken komt dus overeen met de bevinding van Lai [29]: technologie heeft op dit moment vaak een ondersteunende rol en heeft de fundamentele manier van leren niet veranderd.

In verband met aanbevelingen vinden de leerkrachten het belangrijk om controle te krijgen en de leerlingen controle te geven. Tijdens de ontwikkeling van Wiski hebben de leerkrachten de initiële moeilijkheid van de oefeningen gekozen, wat een eerste stap is, maar niet voldoende. Tijdens de gesprekken vertelden ze dat ze graag de mogelijkheid zouden hebben om het niveau van een leerling aan te passen. Het feit dat de leerkrachten graag controle willen geven aan de leerlingen wordt aangemoedigd door de leerlingen zelf. Het controlemechanisme van Wiski werd goed onthaald en was voor het grootste deel van de leerlingen een meerwaarde (cfr. *“Ja. Dit biedt de mogelijkheid om zelf ook een beetje richting te geven aan welke oefeningen die je wilt. Ook als je het wat minder doet, dat je toch ook eens moeilijkere oefeningen krijgt om te zien wat die inhouden.”*). Daarnaast vinden leerkrachten het ook belangrijk dat er meerdere oefeningen tegelijk aanbevolen worden en dat ze verplicht zijn. De leerlingen gaven geen feedback in verband met deze eigenschappen van het systeem.

Tot slot willen leerkrachten graag een dashboard waarmee ze de leerlingen kunnen monitoren. De belangrijkste elementen hiervoor zijn: notificaties met belangrijke inzichten, zowel data over specifieke leerlingen als over specifieke oefeningen, clustering van leerlingen en zo veel mogelijk details.

Uit het onderzoek blijkt dus dat digitale leeromgevingen binnen het onderwijs zeker een toekomst hebben, maar zolang er geen goed alternatief is voor meerkeuzevragen enkel voor remediëring. Het toevoegen van een aanbevelingssysteem waarbij zowel de leerkrachten als de leerlingen controle krijgen, is zeker een meerwaarde. De leerkrachten zouden de controle graag uitvoeren met een dashboard dat daarbovenop enerzijds inzichten geeft over de leerlingen en oefeningen en anderzijds de mogelijkheid biedt om zelf oefeningen toe te voegen.

6.5 Beperkingen van het onderzoek

Het uitgevoerde onderzoek heeft een aantal beperkingen. In deze paragraaf worden ze met hun mogelijke gevolgen besproken.

1. Een eenvoudig Elo-classificatiesysteem is gebruikt als aanbevelingssysteem, waarbij elke gebruiker één niveau heeft. Er bestaan meer geavanceerde algoritmen om aanbevelingen te genereren, bijvoorbeeld het multi-concept multivariaat op Elo gebaseerd leermodel van Abdi et al. [1]. Het model heeft voor elke gebruiker een niveau per onderwerp in plaats van één globaal niveau. Er is niet gekozen voor een meer geavanceerd systeem omdat de focus van het onderzoek niet ligt op het aanbevelingssysteem en er een beperkte implementatietijd was.
2. Het *cold-start* probleem in verband met de moeilijkheden van de oefeningen is deels opgelost doordat leerkrachten zelf de moeilijkheden kozen voor de relevante oefeningen voor hen. Het probleem is dus enkel opgelost voor een minderheid aan oefeningen, de andere hebben een willekeurige startwaarde gekregen. De gebruikers die deelnamen aan het onderzoek hebben enkel oefeningen gemaakt waarvoor leerkrachten de startwaarde hadden gekozen.
3. Binnen dezelfde klasgroep kregen niet alle leerlingen dezelfde versie van de website. Het is dus mogelijk dat bepaalde leerlingen gezien hebben dat er andere versies bestaan, wat de resultaten mogelijk beïnvloed heeft.
4. Zowel de leerkrachten, tijdens de pilootstudie, als de leerlingen, tijdens de *think-aloud* studies, gaven aan dat het belangrijk is om te tonen waarom een oefening fout beantwoord is en dat is niet aanwezig in Wiski.
5. Hoe de leerlingen hun startniveau gekozen hebben, is niet bevraagd omwille van gelimiteerde onderzoekstijd. De mogelijke invloed van het Dunning-Kruger effect kan dus niet onderzocht worden.
6. Een aantal gebruikers geven aan dat ze moeilijk een oordeel kunnen vellen over vertrouwen, omdat ze Wiski nog niet lang genoeg gebruikt hebben. Het vertrouwen van de leerlingen kan dus mogelijk veranderen doorheen de tijd en dat is niet onderzocht. Een aantal citaten die deze bevinding ondersteunen zijn: *“Beetje te weinig ervaring met Wiski om dit direct in te vullen.”*, *“Als je er lang genoeg op oefent misschien wel.”* en *“Ik heb te weinig kunnen oefenen en de site kunnen uitvoeren, dus kan ik ook geen goed eindoordeel geven (op dit moment). Van hetgeen dat ik heb gezien, heeft deze site wel potentieel om een goede manier van oefenen voor wiskunde te worden.”*

7. Ook over het niveau van de leerlingen zijn er opmerkingen omwille van gelimiteerde gebruikstijd. Een aantal leerlingen geven aan dat hun niveau nog niet geconvergeerd is naar het juiste: *“Ik heb nu enkele oefeningen gemaakt en heb nog niet het niveau gevonden dat bij mij past. Dus ik ben meer geneigd om oefeningen in mijn boek te maken omdat ik weet dat we dat niveau moeten kunnen.”* en *“Ja, ik denk zeker dat als ik langer op Wiski oefen, Wiski goede oefeningen voor mij kan aanbevelen.”* Het gevoel van deze leerlingen kan dus mogelijk anders zijn na een langere gebruikstijd.

Hoofdstuk 7

Conclusie

Het doel van het uitgevoerde onderzoek was om het effect te bepalen van controle over de aanbevolen oefeningen. Specifiek is onderzocht of het toevoegen van een controlemechanisme of een controlemechanisme en een uitleg over de impact ervan invloed had op gebruiksgemak, interactie-adequaatheid en vertrouwen in de aanbevelingen. Voor het onderzoek werd het bestaande platform Wiski [41] eerst uitgebreid met een aanbevelingssysteem, een controlemechanisme en uitleg over de impact van de controle door middel van een iteratief gebruikersgestuurd ontwikkelingsproces. Daarna is er een gerandomiseerd onderzoek met controlegroep uitgevoerd met 71 leerlingen om het effect te onderzoeken van enerzijds een controlemechanisme en anderzijds een controlemechanisme en uitleg over de impact ervan.

7.1 Antwoorden op onderzoeksvragen

Onderzoeksvraag 1. Hoe worden waargenomen gebruiksgemak en interactie-adequaatheid beïnvloed door de toevoeging van de extra functionaliteiten?

Waargenomen gebruiksgemak meet de mogelijkheid van de gebruiker om snel en correct taken uit te voeren. Voor geen enkele vergelijking werd een verschil gevonden tussen de groepen. Er kan dus besloten worden dat het toevoegen van controle en eventueel een uitleg over de impact van de controle er niet voor gezorgd hebben dat het moeilijker werd om de website te gebruiken. De resultaten in verband met interactie-adequaatheid en vertrouwen werden dus niet beïnvloed door het waargenomen gebruiksgemak.

Interactie-adequaatheid omvat de integratie van drie interactiemechanismen voorkeursuitlokking, voorkeursherziening en verklaring van resultaten. De resultaten toonden aan dat het toevoegen van controle of het toevoegen van controle en uitleg over de impact ervan interactie-adequaatheid verhogen. Daarnaast bleek ook dat het toevoegen van de uitleg over de impact van de controle ten opzichte van enkel controle de interactie-adequaatheid niet verhoogt. De gebruikers zien de uitleg dus niet als een verklaringen voor de aanbevelingen.

Onderzoeksvraag 2. Hoe wordt het vertrouwen van leerlingen beïnvloed door het geven van controle over de aanbevelingen?

De resultaten toonden niet aan dat het toevoegen van de controle het vertrouwen van de gebruiker verhoogt. Uit de opmerkingen van de leerlingen bleek dat de vraagstelling niet aangepast was aan die van de eigen leerkracht, waardoor een aantal de voorkeur geven aan pen-en-papier oefeningen om evaluatie voor te bereiden. Daarnaast gaven een aantal leerlingen ook aan dat ze de studiekeerperiode te kort vonden om een onderbouwde mening te vormen. Ondanks het feit dat het vertrouwen niet verhoogd is, gaven de meeste leerlingen aan dat de toevoeging van het controlemechanisme een meerwaarde is voor Wiski.

Onderzoeksvraag 3. Hoe wordt het vertrouwen van leerlingen beïnvloed door een uitleg over de impact van de controle van de leerling over de aanbevelingen?

De resultaten tonen dat het toevoegen van controle en een uitleg over de impact ervan de vertrouwensaspecten welwillendheid, eendimensionaal vertrouwen en geloofsovertuiging verhoogt. Aangezien welwillendheid hoger is, geloven de gebruikers dat Wiski om hen geeft en in hun belang handelt. Bij het vergelijken van **CONTROLE** en **CONTROLE+IMPACT** bleken dezelfde vertrouwensaspecten verhoogd, met daarbovenop ook bekwaamheid, wat betekent dat de gebruikers geloven dat Wiski het potentieel, de kunde en de expertise heeft om goede aanbevelingen te leveren.

7.2 Mogelijkheden voor toekomstig onderzoek

Ondanks de vele thema's die het uitgevoerde onderzoek behandelt, is er zeker extra onderzoek mogelijk. Enkele ideeën voor verder onderzoek worden hieronder opgelijst.

- Om het doel van het uitgevoerde onderzoek te bepalen, is er een kleinschalige pilootstudie uitgevoerd met leerkrachten die graag wilden deelnemen aan het onderzoek. Uit die studie kwamen heel wat ideeën naar voren (cfr. Paragraaf 4.1) en een aantal maakten geen deel uit van Wiski, namelijk: alternatieven voor meerkeuzevragen, leerkrachten zelf oefeningen laten toevoegen, mogelijkheid voor leerkrachten om niveau van leerlingen aan te passen, uitgewerkte versies van oplossingen en een dashboard voor leerkrachten om leerlingen te volgen.
- Tijdens de uitgevoerde *think-aloud* studies kwamen de meeste opmerkingen in verband met de uitleg over hoe Wiski werkt. Het kan interessant zijn om te onderzoeken hoe een uitleg over de werking van een algoritme best wordt overgebracht naar kinderen tussen 12 en 18 jaar.
- Bij de implementatie van Wiski is er gekozen om elke leerling één niveau te geven voor alle onderwerpen. Een aanbevelingssysteem ontwikkelen en onderzoeken waarbij elke leerling voor elk onderwerp een specifiek niveau heeft, kan de resultaten uit dit onderzoek ondersteunen of kan tot andere inzichten leiden. Een aantal klassen hebben oefeningen gemaakt over verschillende onderwerpen

en het is mogelijk dat een leerling goed is in bepaalde onderdelen van wiskunde, maar problemen heeft met andere. Als hetzelfde niveau gebruikt wordt voor beide, zal het te hoog zijn voor het ene en te laag voor het andere, wat mogelijk invloed heeft gehad op het vertrouwen in de aanbevelingen. Daarnaast kunnen met meerdere niveaus bijvoorbeeld ook andere vakken toegevoegd worden aan hetzelfde systeem.

- De leerlingen die Wiski gebruikten, moesten tijdens de registratie aangeven wat hun niveau is voor wiskunde. Het effect van de keuze is niet onderzocht, waardoor er geen resultaten waren in verband met het Dunning-Kruger effect.

Bijlage A

Pilootstudie

Dit deel van de appendix toont de vragen en resultaten van de pilootstudie.

A.1 Algemene vragen

TABEL A.1: Vragen gebruikt tijdens pilootstudie.

Vragen uit pilootstudie
1. Welke digitale leeromgevingen ken je?
2. Gebruik je deze ook?
2.1 Wanneer gebruik je deze?
2.2 Wat vind je hier goed en slecht aan?
3. Gebruik je het als summatieve of formatieve evaluatie? [4]
4. Wat vind je van meerkeuzevragen?
5. Zou je graag zelf oefeningen toevoegen aan de applicatie?
6. Als er oefeningen worden aanbevolen aan je leerlingen, zou je die dan graag kunnen sturen?
7. Zou je leerlingen de juiste oplossing van de oefening tonen?
8. Zou het nuttig zijn om leerlingen feedback te laten geven aan het aanbevelingssysteem? [35]
9. Wat heeft het meeste nut:
9.1 De leerling kan zelf kiezen tussen de aanbevolen oefeningen en alle andere oefeningen.
9.2 De leerling moet de aanbevolen oefeningen maken.
10. Heb je liefst één aanbevolen oefening per keer of een reeks aanbevolen oefeningen?

A.2 Reeksen

TABEL A.2: Reeksen gebruikt tijdens pilootstudie [\[36, 45, 47\]](#).

Reeksen uit pilootstudie
1. Noem de belangrijkste kenmerken voor een digitale leeromgeving:
Visuele aantrekkingskracht
Gebruiksvriendelijkheid
Niveau van begrip
Waargenomen nut
Gedragsverandering

A. PILOOTSTUDIE

2. Sorteert naargelang belang voor u (doel applicatie):
 - Meer efficiënt lesgeven.
 - Zekerder worden over het niveau van bepaalde leerlingen.
 - Leerlingen de kans geven om nieuwe oefeningen te laten ontdekken.
 3. Noem de belangrijkste kenmerken voor een digitale leeromgeving:
 - Waargenomen nauwkeurigheid
 - Herkenbaarheid
 - Vernieuwend
 - Aantrekkingskracht
 - Amusant
 - Diversiteit
 - Context compatibiliteit
 4. Belangrijkste drie kenmerken voor aanbevelingen:
 - Transparant
 - Instelbaar
 - Vertrouwen
 - Effectiviteit
 - Overtuigingskracht
 - Efficiënt
 - Tevredenheid
 5. Belangrijkste vier doelen van het dashboard voor leerkrachten:
 - Detecteer studenten in gevaar
 - Monitor gebruik van technologie
 - Timemanagement
 - Detecteer behoefte aan eigen kennis-upgrade
 - Detecteer moeilijkheden binnen het vak
 - Optimaliseer onderwijs-strategie
 - Cluster studenten voor feedback
-

A.3 Resultaten

TABEL A.3: Resultaten pilootstudie (deel 1).

Vraag	Antwoord	Aantal leerkrachten
Gekende platformen	Smartschool	4
	Pelckmans Portaal	2
	Polpo	2
	Bookwidgets	2
	Usolv-it	1
	Diddit	1
	Code.org	1
	Duolingo	1
	Scoodle	1
	Kahoot	1
	Wezooz Academy	1
Idee over meerkeuzevragen	Ja	1
	Neen	3
Summatief of formatief	Summatief	0
	Formatief	4

A.3. Resultaten

Wanneer een digitale leeromgeving gebruiken	In de les	0
	Thuis	4
Gebruik platform	Verplicht	0
	Vrije keuze	4
Leerkracht kan oefeningen toevoegen	Ja	4
	Neen	0
Leerkracht kan aanbevelingen sturen	Ja	3
	Neen	1
Leerling kan aanbevelingen sturen	Ja	3
	Neen	1

TABEL A.4: Resultaten pilootstudie (deel 2).

Vraag (soms worden niet alle punten toegekend)	Antwoord	Punten
Reeks 1 (3, 2 en 1 punt)	Gebruiksvriendelijkheid	11
	Niveau van begrijpen	5
	Waargenomen nut	4
	Gedragsverandering	3
	Visuele aantrekkingskracht	1
Reeks 2 (3, 2 en 1 punt)	Zekerder over niveau	9
	Iets nieuws ontdekken	7
	Efficiënter lesgeven	5
Reeks 3 (3, 2 en 1 punt)	Diversiteit	7
	Herkenbaarheid	6
	Context compatibiliteit	5
	Waargenomen nauwkeurigheid	3
	Amusant	3
	Aantrekkingskracht	2
	Vernieuwend	0
Reeks 4 (3, 2 en 1 punt)	Efficiënt	7
	Transparant	6
	Instelbaar	6
	Vertrouwen	3
	Effectiviteit	1
	Overtuigingskracht	0
	Tevredenheid	0
	Detecteer studenten in gevaar	12
Reeks 5 (4, 3, 2 en 1 punt)	Cluster studenten voor feedback	11
	Detecteer moeilijkheden binnen het vak	9
	Timemanagement	2
	Optimaliseer onderwijs-strategie	2
	Monitor gebruik van technologie	0
	Detecteer behoefte aan eigen kennis-upgrade	0

TABEL A.5: Resultaten pilootstudie (deel 3).

Element	Aantal leerkrachten
Notificaties voor leerkrachten met belangrijke inzichten uit de data.	4/4
Naast data voor specifieke leerling ook data tonen voor specifieke oefening.	3/4
Leerlingen clusteren op basis van hun data.	3/4
Hoe meer details, hoe beter.	3/4
Groeicurve leerling tonen.	1/4
Klassen met elkaar vergelijken.	0/4

Bijlage B

Aanwervingsdocumenten gebruikersstudie

Dit deel van de appendix bevat de documenten die gebruikt zijn voor de aanwerving van deelnemers voor de gebruikersstudie. Het eerste document is de informatiebrochure voor de leerkrachten van de deelnemende leerlingen. Het tweede document is de informatiebrochure voor de leerlingen en hun ouders. Het laatste document is het geïnformeerd toestemmingsformulier voor de leerlingen, en hun ouders als de leerlingen 15 jaar of jonger zijn.

Informatiebrochure Wiski

Beste leerkracht, bedankt voor de interesse in het wetenschappelijk onderzoek voor mijn masterthesis. Hieronder wordt uitgelegd hoe het onderzoek zal verlopen.

Doel van het onderzoek

Er zijn veel verschillen tussen leerlingen. De ene leerling heeft nood aan extra basisoefeningen, de andere aan uitdagende oefeningen. Dit leidt ertoe dat er nood is aan verschillende oefeningen voor verschillende leerlingen op verschillende momenten. Om hier rekening mee te houden heb ik voor mijn thesis een digitaal wiskundeplatform gemaakt met oefeningen van de uitgeverij Die Keure. De website zal dus trachten om de leerling die oefeningen te geven die het beste aansluiten bij zijn/haar niveau. Met mijn onderzoek wil ik te weten komen of de leerling de website vertrouwt.

Hoe verloopt het onderzoek?

Toestemmingsformulier:

Elke leerling moet een toestemmingsformulier invullen voor het deelnemen aan het onderzoek (zie: <https://www.kuleuven.be/english/research/ethics/committees/smec/faq>). Het is hier van uiterst belang dat de leerling in kwestie niet onder druk gezet wordt om deel te nemen aan het onderzoek. Indien de leerling niet wil deelnemen, dient de leerkracht een gelijkwaardig alternatief te voorzien. Dit kunnen bijvoorbeeld oefeningen op papier zijn. Verder kunnen leerlingen op elk moment en zonder reden hun deelname stopzetten, zonder nadeel te ondervinden. Voor leerlingen die 15 jaar of jonger zijn, moet het formulier ingevuld worden door zijn/haar ouders. De leerling zelf moet het formulier ook ondertekenen. Leerlingen van 16 jaar en ouder mogen zelf het formulier invullen.

Gebruik van Wiski:

Om toegang te krijgen tot het platform moet elke leerling zich registreren. Hierbij hoeft de leerling zijn/haar echte naam niet te gebruiken, aangezien alle gegevens anoniem worden bijgehouden. Eenmaal geregistreerd kan de leerling oefeningen beginnen maken. Na de registratie moet de leerling zijn/haar eigen niveau inschatten. Het platform zal deze informatie gebruiken om de juiste oefeningen aan te bieden. Na elke reeks oefeningen zal Wiski het niveau van de leerling aanpassen en een nieuwe reeks oefeningen aanbevelen. Na drie reeksen oefeningen zal de leerling gevraagd worden om een vragenlijst in te vullen die te maken heeft met de ervaringen van de leerling op Wiski. De antwoorden van de leerlingen op deze vragenlijst zijn heel belangrijk voor mijn onderzoek.

Wat gebeurt er met de gegevens?

De gegevens worden veilig bijgehouden en zijn alleen toegankelijk voor mij als verantwoordelijke onderzoeker. Er worden geen persoonlijke gegevens gevraagd die gebruikt kunnen worden om de leerling direct te identificeren. Na het gebruik van de website zal de leerling een vragenlijst moeten invullen over zijn/haar ervaringen op Wiski. Dit gebeurt volledig anoniem en de leerling zal dus, opnieuw, niet identificeerbaar zijn. De anonieme resultaten van het onderzoek worden alleen gebruikt voor wetenschappelijke doeleinden. Alles wordt gepubliceerd in een openbare thesistekst. Na het onderzoek worden de resultaten bijgehouden op een beveiligde server van de KU Leuven.

Is de leerling verplicht om deel te nemen aan het onderzoek?

Neen, de leerling neemt volledig op vrijwillige basis deel aan het onderzoek. Op elk moment kan de leerling beslissen om niet langer deel te nemen. Er hoeft hiervoor ook geen reden geven te worden en er zullen geen nadelen aan verbonden zijn voor de leerling. Indien een leerling niet wenst deel te nemen aan het onderzoek, zal de leerkracht een andere opdracht voorzien (bijvoorbeeld evenwaardige oefeningen op papier).

COVID-19 voorzorgsmaatregelen

Indien de maatregelen het toelaten, zal ik normaal fysiek aanwezig zijn tijdens het onderzoek om eventuele vragen te kunnen beantwoorden en een vlot verloop van het onderzoek te vergemakkelijken. Natuurlijk worden alle voorzorgsmaatregelen van de school nageleefd tijdens het onderzoek.

Wie heeft dit onderzoek gecontroleerd?

Dit onderzoeksproject is goedgekeurd door de Sociaal-Maatschappelijke Ethische Commissie (SMEC) te Leuven. Deze commissie gaat na dat onderzoeken veilig verlopen en dat de rechten van de testpersoon worden gerespecteerd. Voor eventuele klachten of andere bezorgdheden omtrent ethische aspecten van dit onderzoek kan je contact opnemen met SMEC via smec@kuleuven.be.

Heb je nog vragen?

Indien er onduidelijkheden zijn of je hebt nog vragen of opmerkingen, dan kan je steeds contact opnemen met mij via leen.dereu@student.kuleuven.be.

Informatiebrochure Wiski

Beste ouder en leerling, ik ben Leen een masterstudent ingenieurswetenschappen en voor mijn masterthesis voer ik een onderzoek uit in middelbare scholen. De wiskundeleerkracht van uw zoon/dochter is bereid om deel te nemen aan het onderzoek en uw zoon/dochter kan kiezen om deel te nemen. Hieronder zal u lezen wat het onderzoek precies inhoudt.

Doel van het onderzoek

Er zijn veel verschillen tussen leerlingen. De ene leerling heeft nood aan extra basisoefeningen, de andere aan uitdagende oefeningen. Dit leidt ertoe dat er nood is aan verschillende oefeningen voor verschillende leerlingen op verschillende momenten. Dit heb ik voor mijn thesis mogelijk gemaakt via een digitaal wiskundeplatform. Met mijn onderzoek wil ik te weten komen of de leerlingen deze website vertrouwen.

Hoe verloopt het onderzoek?

Toestemmingsformulier: Elke leerling moet een toestemmingsformulier invullen voor het deelnemen aan het onderzoek (zie: <https://www.kuleuven.be/english/research/ethics/committees/smec/faq>). Het is hier van uiterst belang dat de leerling in kwestie niet onder druk gezet wordt om deel te nemen aan het onderzoek. Indien de leerling niet wil deelnemen, zal de leerkracht een gelijkwaardig alternatief voorzien. Verder kunnen leerlingen op elk moment en zonder reden hun deelname stopzetten, zonder nadeel te ondervinden. Voor leerlingen die 15 jaar of jonger zijn, moet het formulier ingevuld worden door de ouders en de leerling zelf. Leerlingen van 16 jaar en ouder mogen zelf het formulier invullen.

Gebruik van Wiski: Om toegang te krijgen tot het platform moet elke leerling zich registreren. Hierbij hoeft de leerling zijn/haar echte naam niet te gebruiken, aangezien alle gegevens anoniem worden bijgehouden. Eenmaal geregistreerd kan de leerling oefeningen beginnen maken. Na elke reeks oefeningen zal Wiski het niveau van de leerling aanpassen en een nieuwe reeks oefeningen aanbevelen. Op het einde van het onderzoek wordt aan de leerling gevraagd om een vragenlijst in te vullen over zijn/haar ervaringen. De antwoorden van de leerlingen op deze vragenlijst zijn heel belangrijk voor mijn onderzoek.

Wat gebeurt er met mijn gegevens?

De gegevens worden veilig bijgehouden en zijn alleen toegankelijk voor mij. Er worden geen persoonlijke gegevens gevraagd die gebruikt kunnen worden om de leerling direct te identificeren. De anonieme resultaten van het onderzoek worden alleen gebruikt voor wetenschappelijke doeleinden en zullen terug te vinden zijn in een openbare thesistekst.

COVID-19 voorzorgsmaatregelen

Indien de maatregelen het toelaten, zal ik normaal fysiek aanwezig zijn tijdens het onderzoek om eventuele vragen te kunnen beantwoorden en een vlot verloop van het onderzoek te vergemakkelijken. Natuurlijk worden alle voorzorgsmaatregelen van de school nageleefd tijdens het onderzoek.

Heb je nog vragen?

Indien er onduidelijkheden zijn of je hebt nog vragen of opmerkingen, dan kan je steeds contact opnemen met mij via leen.dereu@student.kuleuven.be.

Geïnformeerde toestemming

Titel van het onderzoek:

Recommending the right math: adapting digital learning environments to learners and explaining recommendations

Naam + contactgegevens promotor en onderzoeker(s):

Leen Dereu (leen.dereu@student.kuleuven.be),
Jeroen Ooge (jeroen.ooge@kuleuven.be),
Dr. ir. Robin De Croon (robin.decroon@kuleuven.be),
Prof. Dr. Katrien Verbert (katrien.verbert@kuleuven.be)

Doel en methodologie van het onderzoek:

Het doel van het onderzoek is om de nieuwsgierigheid voor en het vertrouwen van de leerling in een digitale leeromgeving te meten. Wiski probeert die oefeningen aan te bevelen die passen bij het niveau van de leerling. De leerling zal eerst vrij oefeningen maken op Wiski en daarna een vragenlijst invullen over zijn of haar ervaringen. Met haar onderzoek wil Leen te weten komen of leerlingen nieuwsgieriger zijn en/of meer vertrouwen hebben in het platform naargelang ze bepaalde varianten van het platform gebruiken.

Duur van het experiment:

Het experiment wordt uitgevoerd in verschillende scholen op verschillende tijdstippen. De leerling zal ongeveer een lesuur besteden aan het experiment.

- Ik begrijp wat van mij verwacht wordt tijdens dit onderzoek.
- Ik weet dat ik zal deelnemen aan volgende proeven of testen:
 - Maken van oefeningen op Wiski.
 - Beantwoorden van vragenlijsten.
- Ik weet dat er risico's of ongemakken kunnen verbonden zijn aan mijn deelname:
 - Niet van toepassing.
- Ikzelf of anderen kunnen baat bij dit onderzoek hebben op volgende wijze:
 - Je kan tijdens het onderzoek gebruik maken van Wiski dat kwalitatieve oefeningen bevat die normaal enkel te maken zijn op een betalend alternatief (namelijk Polpo van Die Keure). Je kan bovendien je wiskundige kennis uitbreiden zonder prestatiedruk.
- Ik weet dat er een beloning of compensatie gekoppeld is aan mijn deelname aan het onderzoek:
 - Gebruik maken van oefeningen gegeven door uitgeverij Die Keure.
- Ik begrijp dat mijn deelname aan deze studie vrijwillig is. Ik heb het recht om mijn deelname op elk moment stop te zetten. Daarvoor hoef ik geen reden te geven en ik weet dat daaruit geen nadeel voor mij kan ontstaan.

Voor de verdere verwerking van de verzamelde gegevens geldt het algemeen belang als rechtsgrond volgens de AVG/GDPR. Stopzetting van deelname aan de studie houdt dus in dat de eerder verzamelde gegevens nog verder rechtsgeldig kunnen worden betrokken in de studie en niet moeten worden verwijderd door KU Leuven.
- De resultaten van dit onderzoek kunnen gebruikt worden voor wetenschappelijke doeleinden en mogen gepubliceerd worden. Mijn naam wordt daarbij niet gepubliceerd, anonimiteit en de vertrouwelijkheid van de gegevens is in elk stadium van het onderzoek gewaarborgd.
- Ik wil graag op de hoogte gehouden worden van de resultaten van dit onderzoek. De onderzoeker mag mij hiervoor contacteren op het volgende e-mailadres:

.....

Opgemaakt in tweevoud.

- Voor vragen evenals voor de uitoefening van mijn rechten (inzage gegevens, correctie ervan,...) weet ik dat ik na mijn deelname terecht kan bij:
Leen Dereu (leen.dereu@student.kuleuven.be)

Meer informatie met betrekking tot privacy in onderzoek kan ik terugvinden op www.kuleuven.be/privacy. Verdere vragen over privacyaspecten kan ik richten tot de data protection officer: dpo@kuleuven.be
- Voor eventuele klachten of andere bezorgdheden omtrent ethische aspecten van deze studie kan ik contact opnemen met de Sociaal-Maatschappelijke Ethische Commissie van KU Leuven: smec@kuleuven.be
- Ik weet dat ik bij onderstaande terecht kan indien ik na het onderzoek ongemakken of moeilijkheden ervaar als gevolg van de thema's die in het onderzoek aan bod kwamen:
Leen Dereu (leen.dereu@student.kuleuven.be)

Ik heb bovenstaande informatie gelezen en begrepen en heb antwoord gekregen op al mijn vragen betreffende deze studie. Ik stem toe om deel te nemen.

Belangrijk: Als je 15 jaar of jonger bent, dan moeten je ouders dit document verplicht ondertekenen.

Datum:

Naam en handtekening proefpersoon

Naam en handtekening onderzoeker



Naam en handtekening(en) ouder(s) **(verplicht voor leerlingen die 15 jaar of jonger zijn)**

Bijlage C

Think-aloud studies

Dit deel van de appendix toont welke vragen en taken gebruikt zijn tijdens de *think-aloud* studies. De vragen zijn opgedeeld in twee fases. De eerste focust op de navigatie doorheen het prototype en de tweede is een algemene bespreking na het doorlopen van het prototype.

C.1 Eerste *think-aloud* studie

C.1.1 Fase 1: Navigatie

TABEL C.1: De vragen en taken uit de eerste fase (navigatie) van de eerste *think-aloud* studie.

Vragen en taken uit de eerste fase van de eerste <i>think-aloud</i> studie
1. Registreer jezelf op deze website.
2. Maak oefeningen over “Hoofdbewerkingen”.
3. Pagina “Hoe goed denk je dat je bent...”
3.1 Wat betekenen de verschillende lampjes volgens jou?
3.2 Wat vind je van de keuze voor lampjes?
3.3 Wat vind je van deze pagina in het algemeen?
4. Maak de eerste oefening en geef hierna aan hoe je het niveau van de oefening vond.
5. Maak de rest van de oefeningen. Antwoord juist en fout.
6. Pagina “Wiski wil graag extra informatie...”
6.1 Wat is de bedoeling van deze pagina volgens jou?
6.2 Wat vind je van deze pagina?
7. Pagina “Hoe is je nieuw niveau bepaald?”
7.1 Hoe kan je meer informatie krijgen over hoe Wiski werkt?
7.2 Hoe werkt Wiski volgens jou?
7.3 Ga naar de volgende pagina met uitleg.
7.4 Hoe kan je Wiski helpen volgens jou?
7.5 Wat vind je van deze pagina’s?
7.6 Zou je tijdens het gebruik van Wiski willen weten hoe het werkt? Vind je deze uitleg nuttig?
7.7 Sluit de uitleg af.

8. Pagina met lijngrafiek als visualisatie.
 - 8.1 Wat wordt er op de grafiek getoond volgens jou?
 - 8.2 Hoe vind je de animatie?
 - 8.3 Wat vind je van deze pagina?
 9. Maak de oefening en onderbreek daarna de reeks oefeningen.
 10. Bezoek je profiel.
 11. Hoe kan je uitleg krijgen over de kleur van de lamp?
 - 11.1 Wat vind je van deze uitleg?
 12. Hoe kan je je profiel bewerken?
-

C.1.2 Fase 2: Algemene bespreking

TABEL C.2: De vragen en taken uit de tweede fase (algemene bespreking) van de eerste *think-aloud* studie.

Vragen en taken uit de tweede fase van de eerste *think-aloud* studie

1. Er zijn ook nog twee andere mogelijkheden voor verklaring van het systeem over je niveau.
 - 1.1 Kan je me uitleggen wat zichtbaar is op de grafiek?
 - 1.2 Hoe vind je de animatie?
 - 1.3 Wat vind je van deze pagina?
 2. Welke van de drie soorten zou jij het liefste zien?
 3. Kan je me uitleggen wat de bedoeling was van deze website?
 4. Vragenlijst: Kan je de vragenlijst voor mij doorlopen? Ik weet dat het niet mogelijk is om alle vragen te beantwoorden na het bekijken van een prototype. Je hoeft de vraag dus enkel expliciet te beantwoorden als je effectief een antwoord hebt. Het is voor mij vooral belangrijk om te weten dat je de vragen allemaal begrijpt.
-

C.2 Tweede *think-aloud* studie

C.2.1 Fase 1: Navigatie

TABEL C.3: De vragen en taken uit de eerste fase (navigatie) van de tweede *think-aloud* studie.

Vragen en taken uit de eerste fase van de tweede *think-aloud* studie

1. Registreer jezelf op deze website.
2. Pagina “Hoe goed denk je dat je bent...”
 - 2.1 Wat betekenen de verschillende lampjes volgens jou?
 - 2.2 Wat vind je van de keuze voor lampjes?
 - 2.3 Wat vind je van de keuze van de kleuren van de lampjes?
 - 2.4 Wat vind je van deze pagina in het algemeen?
 - 2.5 Geen een inschatting van je niveau.

3. Hoe werkt Wiski?
 - 3.1 Je wil graag extra informatie, hoe doe je dat?
 - 3.2 Hoe werkt Wiski volgens jou?
 - 3.3 Wat vind je van deze pagina?
 - 3.4 Hoe kan je Wiski helpen volgens jou?
 - 3.5 Wat vind je van deze pagina?
 - 3.6 Zou je tijdens het gebruik van Wiski willen weten hoe het werkt? Vind je deze uitleg nuttig?
 - 3.7 Zou je graag weten waar de initiële moeilijkheid van de oefeningen vandaan komt? Indien ja, hoe wil je de info krijgen?
 - 3.8 Sluit de uitleg af.
 4. Profiel
 - 4.1 Wat zie je op deze pagina?
 - 4.2 Hoe kan je meer uitleg krijgen over de kleur van de lamp en je niveau?
 - 4.3 Ga naar de pagina waar je kan oefenen.
 5. Maak oefeningen over “Volgorde van bewerkingen” van “Natuurlijke getallen”.
 6. Maak de oefeningen en antwoord zowel juist als fout.
 7. Wat vond je van deze pagina’s?
 8. “Wiski wil graag extra informatie...”
 - 8.1 Wat is de bedoeling van deze pagina volgens jou?
 - 8.2 Wat vind je van deze pagina?
 9. Dien je feedback in.
 10. “Hoe is je nieuw niveau bepaald?”
 - 10.1 Bekijk de uitleg.
 - 10.2 Wat wordt er op de grafiek getoond volgens jou?
 - 10.3 Hoe vind je de animatie?
 - 10.4 Wat vind je van deze pagina?
 - 10.5 Maak nieuwe oefeningen over dit onderwerp.
 11. Onderbreek het oefenen.
-

C.2.2 Fase 2: Algemene bespreking

TABEL C.4: De vragen en taken uit de tweede fase (algemene bespreking) van de tweede *think-aloud* studie.

Vragen en taken uit de tweede fase van de tweede *think-aloud* studie

1. Kan je mij uitleggen wat de bedoeling is van deze website?
 2. Ik heb nog een alternatief voor de lampjes.
 - 2.1 Wat vind je van deze keuze?
 - 2.2 Welke visualisatie heeft je voorkeur? En waarom?
 4. Vragenlijst: Kan je de vragenlijst voor mij doorlopen? Ik weet dat het niet mogelijk is om alle vragen te beantwoorden na het bekijken van een prototype. Wat je juist antwoord op de vragen is dus niet van belang. Het is voor mij vooral belangrijk om te weten dat je de vragen allemaal begrijpt.
-

Bijlage D

Vragenlijst

Dit deel van de appendix bevat de vragenlijst van de gebruikersstudie.

D.1 Demografie

TABEL D.1: Deze vragen gaan dieper in op de demografie van de gebruiker. De mogelijke antwoorden staan in de tweede kolom.

Vraag	Mogelijke antwoorden
Leeftijd	Jonger dan 11 - 11 - 12 - 13 - 14 - 15 - 16 - 17 - 18 - Ouder dan 18
Jaar	Lager onderwijs - 1ste middelbaar - 2de middelbaar - 3de middelbaar - 4de middelbaar - 5de middelbaar - 6de middelbaar - Hoger onderwijs
Hoeveel uur wiskundeles?	Minder dan 2 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - Meer dan 8
Hoe graag doe je wiskunde?	Helemaal niet graag - Niet graag - Eerder niet graag - Neutraal - Eerder graag - Graag - Heel Graag

D.2 Vertrouwen in de aanbevelingen

TABEL D.2: Deze vragen gaan dieper in op het vertrouwen van de gebruiker in Wiski. Alle vragen behalve de laatste worden beantwoord met een 7-punt Likertschaal.

Nr.	Nederlandse vertaling	Engels origineel
	Bekwaamheid	Competence
V1	Wiski is zoals een expert (bv. een leerkracht) in wiskunde-oefeningen aanraden.	Wiski is like an expert (for example, a teacher) for recommending math exercises.
V2	Wiski heeft de expertise (kennis) om mijn wiskundeniveau te kunnen inschatten.	Wiski has the expertise (knowledge) to estimate my math level.
V3	Wiski kan mijn wiskundeniveau inschatten.	Wiski can estimate my math level.
V4	Wiski begrijpt de moeilijkheidsgraad van wiskunde-oefeningen goed.	Wiski understands the difficulty level of math exercises well.

D. VRAGENLIJST

V5	Wiski houdt rekening met mijn wiskundenniveau om oefeningen aan te raden.	Wiski takes my math level into account when recommending exercises.
	Welwillendheid	Benevolence
V6	Wiski zet op de eerste plaats dat ik vorderingen maak in wiskunde.	Wiski prioritizes that I improve in math.
V7	Wanneer Wiski oefeningen aanraadt, doet Wiski dat zodat ik vorderingen maak in wiskunde.	Wiski recommends exercises so that I improve in math.
V8	Wiski wil mijn wiskundenniveau goed inschatten.	Wiski wants to estimate my math level well.
	Integriteit	Integrity
V9	Wiski raadt oefeningen op een zo correct mogelijke manier aan.	Wiski recommends exercises as correctly as possible.
V10	Wiski is eerlijk.	Wiski is honest.
V11	Wiski maakt oprechte aanbevelingen.	Wiski makes integrous recommendations.
	Vertrouwen (eendimensionaal)	Trust (one-dimensional)
V12	Ik vertrouw Wiski om mij wiskunde-oefeningen aan te raden.	I trust Wiski to recommend me math exercises.
	Intentie om terug te keren	Intention to return
V13	Als ik nog eens online wiskunde-oefeningen maak, dan kies ik voor Wiski.	If I want to solve math exercises again, I will choose Wiski.
V14	Als ik nog eens wiskunde-oefeningen aangeraden wil krijgen, dan kies ik voor Wiski.	If I want to be recommended math exercises again, I will choose Wiski.
	Open vraag	
	Heb jij vertrouwen in Wiski voor het aanbevelen van oefeningen? Waarom wel/niet?	

D.3 Controle en transparantie

TABEL D.3: Deze vragen gaan dieper in op het gevoel van de gebruiker tegenover de controle die Wiski geeft en hoe transparant Wiski is. Alle vragen behalve de laatste worden beantwoord met een 7-punt Likertschaal. De laatste vraag is een open vraag en is optioneel.

Nr.	Nederlandse vertaling	Engels origineel
	Controle en transparantie	Control and transparency
V15	Ik heb het gevoel dat ik Wiski kan vertellen wat ik wil.	I feel in control of telling the recommender what I want.
V16	Ik heb niet het gevoel dat ik Wiski kan vertellen wat ik wil.	I don't feel in control of telling the system what I want.
V17	Ik heb niet het gevoel dat ik controle heb over het omschrijven en veranderen van mijn voorkeuren.	I don't feel in control of specifying and changing my preferences.
V18	Ik begreep waarom de oefeningen aan mij werden aanbevolen.	I understood why the items were recommended to me.

V19	Wiski helpt mij te begrijpen waarom de oefeningen aan mij werden aanbevolen.	The system helps me understand why the items were recommended to me.
V20	Wiski lijkt mijn keuzeproces te controleren in plaats van ikzelf.	The system seems to control my decision process rather than me.

Open vraag

Heb je nog opmerkingen bij je antwoorden?

D.4 interactie-adequaatheid en waargenomen gebruiksgemak

TABEL D.4: Deze vragen gaan dieper in op het gevoel van de gebruiker tegenover waargenomen gebruiksgemak en interactie-adequaatheid. Alle vragen behalve de laatste worden beantwoord met een 7-punt Likertschaal. De laatste vraag is een open vraag en is optioneel.

Nr.	Nederlandse vertaling	Engels origineel
	Interactie-adequaatheid	Interaction adequacy
V21	Wiski laat me op een geschikte manier mijn voorkeuren aangeven.	The recommender provides an adequate way for me to express my preferences.
V22	Wiski laat me op een geschikte manier mijn voorkeuren aanpassen.	The recommender provides an adequate way for me to revise my preferences.
V23	Wiski legt uit waarom de oefeningen aan mij worden aanbevolen.	The recommender explains why the products are recommended to me.
	Voorkeurspeiling	Preference elicitation
V24	Ik vond het gemakkelijk om Wiski over mijn voorkeuren te vertellen.	I found it easy to tell the system about my preferences.
V25	Het is gemakkelijk om te leren hoe ik Wiski kan vertellen wat ik leuk vind.	It is easy to learn to tell the system what I like.
V26	Het kostte te veel moeite om Wiski te vertellen wat ik leuk vind.	It required too much effort to tell the system what I like.
	Voorkeursherziening	Preference revision
V27	Ik vond het gemakkelijk om Wiski mij verschillende dingen te laten aanbevelen.	I found it easy to make the system recommend different things to me.
V28	Het is gemakkelijk om Wiski te leren mijn voorkeuren aan te passen.	It is easy to train the system to update my preferences.
V29	Ik vond het gemakkelijk om de aanbevolen oefeningen te wijzigen met mijn voorkeursveranderingen.	I found it easy to alter the outcome of the recommended items due to my preference changes.
V30	Het is voor mij gemakkelijk om Wiski te laten weten of ik de aanbevolen oefeningen leuk/niet leuk vind.	It is easy for me to inform the system if I dislike/like the recommended item.
V31	Het is voor mij gemakkelijk om een nieuwe reeks aanbevolen oefeningen te krijgen.	It is easy for me to get a new set of recommendations.
	Open vraag	
	Heb je nog opmerkingen bij je antwoorden?	

D.5 Controle en uitleg over impact ervan

TABEL D.5: Deze vragen gaan dieper in op controle over de aanbevelingen en een uitleg over de impact ervan. Het zijn beide open vragen en ze zijn optioneel.

Vraag

GEEN

Zou je het leuk vinden om na een reeks oefeningen feedback te kunnen geven aan Wiski? Dit kan dan bijvoorbeeld zoals in onderstaande afbeelding.

Zou je het leuk vinden om een verklaring te krijgen van Wiski over wat er gebeurd is met je niveau? Dit kan er dan bijvoorbeeld uitzien zoals in onderstaande afbeelding.

CONTROLE

Wil je nog iets vertellen over de mogelijkheid die je had om feedback te geven aan Wiski? De afbeelding hieronder toont hoe je dat kon doen.

Zou je het leuk vinden om een verklaring te krijgen van Wiski over wat er gebeurd is met je niveau? Dit kan er dan bijvoorbeeld uitzien zoals in onderstaande afbeelding.

CONTROLE+IMPACT

Wil je nog iets vertellen over de mogelijkheid die je had om feedback te geven aan Wiski? De afbeelding hieronder toont hoe je dat kon doen.

Wil je nog iets vertellen over de verklaringen die Wiski gaf? De afbeelding hieronder toont een voorbeeld van een verklaring.

Bijlage E

Initiële moeilijkheid van oefeningen

Dit deel van de appendix bevat de oefeningen met bijhorende initiële moeilijkheid bepaald door de deelnemende leerkrachten. De eerste en tweede kolom zijn de *node-id* en *version-id* en op basis daarvan kunnen specifieke oefeningen geïdentificeerd worden. De derde kolom bevat de moeilijkheid van de oefening: beginner - 1100, gevorderde beginner - 1300, competent - 1500, bedreven - 1700 en expert - 1900.

TABEL E.1: Veeltermuitdrukkingen
(Veeltermen - 3de en 4de middelbaar)

nid	vid	Moeilijkheid
5448	5485	1100
5440	5477	1300
5449	5486	1100
5441	5478	1300
5450	5487	1500
5442	5479	1700
5443	5480	1700
5444	5481	1700
5436	5473	1900
5445	5482	1100
5437	5474	1900
5446	5483	1100
5438	5475	1900
5447	5484	1100
5439	5476	1900

TABEL E.2: Rekenen met logaritmische functies (Exponentiële en logaritmische functies - 5de en 6de middelbaar)

nid	vid	Moeilijkheid
6543	6580	1300
6485	6522	1500
6544	6581	1100
6545	6582	1500
6536	6573	1500
6537	6574	1700
6538	6575	1700
6546	6583	1900
6539	6576	1300
6547	6584	1700
6540	6577	1300
6548	6585	1500
6541	6578	1700
6549	6586	1900
6542	6579	1100

TABEL E.3: Machten (Gehele getallen - Basiskennis)

nid	vid	Moeilijkheid
5044	5081	1900
5036	5073	1500
5045	5082	1300
5037	5074	1900
5046	5083	1500
5038	5075	1900
5047	5084	1900
5039	5076	1500
5040	5077	1700
5032	5069	1300
5049	5086	1300
5041	5078	1700
5033	5070	1500
5048	7473	1900
5042	5079	1700
5034	5071	1500
5043	5080	1900
5035	5072	1500

TABEL E.4: Eerstegraadsvergelijkingen (Gehele getallen - Basiskennis)

nid	vid	Moeilijkheid
4982	5019	1100
4967	5004	1300
4998	5035	1900
4954	4991	1700
4983	5020	1300
4968	5005	1300
4999	5036	1900
4955	4992	1700
4985	5022	1100
4970	5007	1100
5002	5039	1700
4957	4994	1900
4989	5026	1300
4971	5008	1100
5004	5041	1700
4959	4996	1100
5005	5042	1700
4961	4998	1300
4991	5028	1300
4942	4979	1100
4973	5010	1300
5006	5043	1100
4962	4999	1100
4992	5029	1300
4945	4982	1300
4975	5012	1300
5009	5046	1300
4964	5001	1100
4993	5030	1100
4949	4986	1500
4978	5015	1100
4965	5002	1300
4997	5034	1900
4952	4989	1700
4981	5018	1100

TABEL E.5: Elementaire combinatoriek (Telproblemen - 3de middelbaar en 4de middelbaar)

nid	vid	Moeilijkheid
7600	11017	1500
7601	11018	1500
7602	11019	1900
7603	11020	1500
7604	11021	1900
7605	11022	1900
7606	11023	1500
7607	11024	1700
7608	11025	1500
7609	11026	1700
7610	11027	1500
7611	11028	1700
7616	11034	1300
7615	11033	1300
7614	11032	1100
7613	11031	1100
7612	11030	1100

TABEL E.6: Rekenen met lettervormen (Gehele getallen - Basiskennis)

nid	vid	Moeilijkheid
4907	4944	1100
4897	4934	1700
4908	4945	1300
4898	4935	1500
4910	4947	1500
4900	4937	1300
4911	4948	1500
4901	4938	1700
4902	7471	1900
4903	4940	1500
4909	7472	1100
4904	4941	1700
4899	8011	1900
4905	4942	1300
4906	4943	1100
4896	4933	1500

TABEL E.7: Volgorde van bewerkingen (Natuurlijke getallen - Basiskennis)

nid	vid	Moeilijkheid
4887	4924	1300
4879	4916	1700
4871	4908	1900
4888	4925	1500
4880	4917	1500
4872	4909	1900
4873	4910	1900
4889	4926	1700
4881	4918	1500
4874	4911	1300
4882	4919	1500
4875	4912	1700
4883	4920	1700
4876	4913	1500
4884	4921	1500
4885	4922	1500
4877	4914	1700
4869	4906	1100
4886	4923	1300
4878	4915	1700
4870	4907	1900

Bibliografie

- [1] S. Abdi, H. Khosravi, S. Sadiq, and D. Gasevic. A multivariate elo-based learner model for adaptive educational systems, 2019.
- [2] S. Abdi, H. Khosravi, S. Sadiq, and D. Gasevic. Complementing educational recommender systems with open learner models. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, LAK'20*, pages 360–365, New York, NY, USA, 2020. Association for Computing Machinery.
- [3] A. Adadi and M. Berrada. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, 6:52138–52160, 2018.
- [4] W. Admiraal, J. Vermeulen, and J. Bulterman-Bos. Teaching with learning analytics: how to connect computer-based assessment data with classroom instruction? *Technology, Pedagogy and Education*, 29(5):577–591, 2020.
- [5] L. Anderson, D. Krathwohl, P. Airasian, K. Cruikshank, R. Mayer, P. Pintrich, J. Raths, and M. Wittrock. *Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Pearson, 2000.
- [6] K. Boggess, S. Chen, and L. Feng. Towards personalized explanation of robot path planning via user feedback, 2021.
- [7] S. Bull and J. Kay. *Open Learner Models*, volume 308, pages 301–322. Springer Berlin Heidelberg, 09 2010.
- [8] R. Burke. Hybrid web recommender systems. In *The Adaptive Web: Methods and Strategies of Web Personalization*, volume 4321, 01 2007.
- [9] C.-M. Chen, H.-M. Lee, and Y.-H. Chen. Personalized e-learning system using item response theory. *Computers & Education*, 44(3):237–255, 2005.
- [10] C. Conati, O. Barral, V. Putnam, and L. Rieger. Toward personalized xai: A case study in intelligent tutoring systems. *Artificial Intelligence*, 298:103–503, 2021.
- [11] L. Corrin and P. de Barba. Exploring students' interpretation of feedback delivered through learning analytics dashboards. *Rhetoric and Reality: Critical perspectives on educational technology. Proceedings ascilite Dunedin*, pages 629–633, 11 2014.

- [12] J. Dabolins and J. Grundspenkis. The role of feedback in intelligent tutoring system. *Applied Computer Systems*, 14:88–93, 06 2013.
- [13] B. Davis, M. Glenski, W. Sealy, and D. Arendt. Measure utility, gain trust: Practical advice for xai researcher, 2020.
- [14] H. Drachsler, K. Verbert, O. C. Santos, and N. Manouselis. *Panorama of Recommender Systems to Support Learning*, pages 421–451. Springer US, Boston, MA, 2015.
- [15] S. E. Dreyfus. The five-stage model of adult skill acquisition. *Bulletin of Science, Technology & Society*, 24(3):177–181, 2004.
- [16] D. Dunning. Chapter five: The dunning-kruger effect: On being ignorant of one’s own ignorance. In J. M. Olson and M. P. Zanna, editors, *Advances in Experimental Social Psychology*, volume 44, pages 247–296. Academic Press, 2011.
- [17] T. J. H. M. Eggen and A. J. Verschoor. Optimal testing with easy or difficult items in computerized adaptive testing. *Applied Psychological Measurement*, 30(5):379–393, 2006.
- [18] A. E. Elo. *The Rating of Chess Players, Past & Present*. Arco Publishing, INC., 2 edition, 1978.
- [19] J. Fang, Q. Guo, H. Chang, M. Yang, W. Zhao, and Y. Qin. Knowledge recommendation based on item response theory. In *2019 15th International Conference on Semantics, Knowledge and Grids (SKG)*, pages 1–7, 2019.
- [20] G. R. C. for Artificial Intelligence. Interactive machine learning. <https://www.dfki.de/en/web/research/research-departments/interactive-machine-learning>, 2022. Gelezen op: 29-05-2022.
- [21] S. Garcia-Martinez and A. Hamou-Lhadj. *Educational Recommender Systems: A Pedagogical-Focused Perspective*, pages 113–124. Springer International Publishing, Heidelberg, 2013.
- [22] M. Hind. Explaining explainable AI. *XRDS: Crossroads, The ACM Magazine for Students*, 25(3):16–19, 2019.
- [23] R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman. Metrics for explainable ai: Challenges and prospects, 2019.
- [24] D. T. Kadengye, E. Ceulemans, and W. Van Den Noortgate. Modeling growth in electronic learning environments using a longitudinal random item response model. *The Journal of Experimental Education*, 83(2):175–202, 2015.

-
- [25] Katholiek Onderwijs Vlaanderen. Algemene inleiding leerplannen secundair onderwijs. <https://pro.katholiekonderwijs.vlaanderen/katholieke-dialoogschool/basisteksten/algemene-inleiding-leerplannen-secundair-onderwijs>, 2018. Gelezen op: 26-10-2021.
- [26] S. Kato. Practicing the right math: Enhancing trust in an e-learning platform using an explainable recommender system, 2021.
- [27] S. Klinkenberg, M. Straatemeier, and H. L. J. van der Maas. Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, 57(2):1813–1824, 2011.
- [28] U.-M. Krause, R. Stark, and H. Mandl. The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction*, 19(2):158–170, 2009.
- [29] K.-W. Lai. Digital technology and the culture of teaching and learning in higher education. *Australasian Journal of Educational Technology*, 27(8):1263–1275, 2011.
- [30] B. L. Leech. Asking questions: Techniques for semistructured interviews. *PS: Political Science & Politics*, 35(4):665–668, 2002.
- [31] Leerling 2020. Opdracht feedback, feed up, feed forward. <https://leerling2020.nl/wp-content/uploads/2018/12/Opdracht-feedback-geven-Reinier-Geurts.pdf>, 2018. Gelezen op: 28-10-2021.
- [32] N. N. Liu, E. W. Xiang, M. Zhao, and Q. Yang. Unifying explicit and implicit feedback for collaborative filtering. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pages 1445–1448, New York, NY, USA, 2010. Association for Computing Machinery.
- [33] A. Mabbott and S. Bull. Student preferences for editing, persuading, and negotiating the open learner model. In *Intelligent Tutoring Systems*, pages 481–490, 06 2006.
- [34] N. Manouselis, H. Drachsler, K. Verbert, and E. duval. *Recommender Systems for Learning*. Springer New York, NY, 04 2012.
- [35] P. Michlík and M. Bieliková. Exercises recommending for limited time learning. *Procedia Computer Science*, 1(2):2821–2828, 2010.
- [36] M. Millecamp, T. Broos, T. De Laet, and K. Verbert. DIY: learning analytics dashboards. *Companion Proceeding of the 9th International Conference on Learning Analytics & Knowledge (LAK'19)*, pages 947–954, 2019.

- [37] T.-N. Nguyen, L. Drumond, A. Krohn-Grimberghe, and L. Schmidt-Thieme. Recommender system for predicting student performance. *Procedia Computer Science*, 1(2):2811–2819, 2010.
- [38] Onderwijs van de Vlaamse Gemeenschap. De taxonomie van bloom uit de doeken. <https://pro.g-o.be/nieuws/5146>, 2021. Gelezen op: 26-10-2021.
- [39] Onderwijs van de Vlaamse Gemeenschap. Leerplan A-stroom: basisvorming. <https://pro.g-o.be/pedagogische-begeleiding-leerplannen-nascholing/leerplannen/leerplannen-so/nieuwe-leerplannen>, 2021. Gelezen op: 26-10-2021.
- [40] Onderwijs Vlaanderen. Algemene uitgangspunten A-stroom eerste graad. <https://onderwijsdoelen.be/uitgangspunten/4647>, 2018. Gelezen op: 26-10-2021.
- [41] J. Ooge. Het personaliseren van motivationele strategieën en gamificationstechnieken m.b.v. recommendersystemen, 2019.
- [42] J. Ooge, G. Stiglic, and K. Verbert. Explaining artificial intelligence with visual analytics in healthcare. *WIREs Data Mining and Knowledge Discovery*, 2021.
- [43] J. Ooge and K. Verbert. Trust in prediction models: a mixed-methods pilot study on the impact of domain expertise. In *2021 IEEE Workshop on TRust and EXpertise in Visual Analytics (TRES)*, pages 8–13, 09 2021.
- [44] J. Papousek, R. Pelánek, and V. Stanislav. Adaptive practice of facts in domains with varied prior knowledge. In *EDM*, 2014.
- [45] Y. Park and I.-H. Jo. Factors that affect the success of learning analytics dashboards. *Educational Technology Research and Development*, 67:1547–1571, 2019.
- [46] R. Pelánek. Applications of the Elo rating system in adaptive educational systems. *Computers & Education*, 98:169–179, 2016.
- [47] P. Pu and L. Chen. A user-centric evaluation framework of recommender systems. *Proceedings of the fifth ACM conference on Recommender Systems*, 2011.
- [48] F. Ricci, L. Rokach, and B. Shapira. *Recommender Systems Handbook*, pages 1–35. Springer New York, NY, 10 2010.
- [49] C. Schumacher and D. Ifenthaler. Features students really expect from learning analytics. *Computers in Human Behavior*, 78:397–407, 2018.
- [50] Scribbr. T-test begrijpen, uitvoeren (spss) en het resultaat interpreteren. <https://www.scribbr.nl/statistiek/t-toets/>, 2021. Laatst gelezen op: 29-05-2022.

-
- [51] G. Sedrakyan, J. Malmberg, K. Verbert, S. Järvelä, and P. A. Kirschner. Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, 107:105512, 2020.
- [52] K. Sokol and P. Flach. One explanation does not fit all. *Künstliche Intelligenz*, 34:235–250, 2020.
- [53] T. van der Veen & J. van der Wal. Zelfregulerend leren. <https://www.ensie.nl/van-leertheorie-naar-onderwijspraktijk/zelfregulerend-leren>, 2017. Gelezen op: 14-12-2021.
- [54] D. T. Tempelaar, B. Rienties, and B. Giesbers. In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, 47:157–167, 2015. Learning Analytics, Educational Data Mining and data-driven Educational Decision Making.
- [55] S. Teso and K. Kersting. Explanatory interactive machine learning. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '19, page 239–245, New York, NY, USA, 2019. Association for Computing Machinery.
- [56] S. tools for high-throughput data analysis. Unpaired two-samples t-test in r. <http://www.sthda.com/english/wiki/unpaired-two-samples-t-test-in-r>, 2016. Laatst gelezen op: 29-05-2022.
- [57] Toolshero. Dreyfus model voor vaardigheidsverwerving. <https://www.toolshero.nl/human-resources/dreyfus-model-voor-vaardigheidsverwerving/>, 2019. Gelezen op: 1-11-2021.
- [58] Universiteit van Sint Andrews. Analysing likert scale/type data. <https://www.st-andrews.ac.uk/media/ceed/students/mathssupport/Likert.pdf>, 2014. Gelezen op: 03-05-2022.
- [59] K. Wauters, P. Desmet, and W. Van Den Noortgate. Adaptive item-based learning environments based on the item response theory: possibilities and challenges. *Journal of Computer Assisted Learning*, 26(6):549–562, 2010.
- [60] B. J. Zimmerman. Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1):3–17, 1990.