Faculteit Industriële Ingenieurswetenschappen

master in de industriële wetenschappen: informatica

Masterthesis

Geautomatiseerde en gedigitaliseerde inschatting van competenties

Lars Gielen

Scriptie ingediend tot het behalen van de graad van master in de industriële wetenschappen: informatica

PROMOTOR :
Prof. dr. Maarten WIJNANTS
dr. Eva GEURTS

PROMOTOR :
Dhr. Benny CLAES

COPROMOTOR :
De heer Arno VERSTRAETE

Gezamenlijke opleiding UHasselt en KU Leuven

Universiteit Hasselt | Campus Diepenbeek | Faculteit Industriële Ingenieurswetenschappen | Agoralaan Gebouw H - Gebouw B | BE 3590 Diepenbeek

Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE 3590 Diepenbeek
Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE 3500 Hasselt

2024
2025

Faculteit Industriële Ingenieurswetenschappen

master in de industriële wetenschappen: informatica

Masterthesis

Geautomatiseerde en gedigitaliseerde inschatting van competenties

Lars Gielen

Scriptie ingediend tot het behalen van de graad van master in de industriële wetenschappen: informatica

PROMOTOR :
Prof. dr. Maarten WIJNANTS
dr. Eva GEURTS

PROMOTOR :
Dhr. Benny CLAES

COPROMOTOR :
De heer Arno VERSTRAETE

UHASSELT KU LEUVEN

Woord vooraf

Allereerst gaat mijn oprechte dank uit naar mijn promotor, Prof. dr. Maarten Wijnants, en co-promotor, dr. Eva Geurts. Hun deskundige begeleiding en kritische inzichten waren van onschatbare waarde. Ook wil ik mijn begeleider, Arno Verstraete, bedanken voor de goede begeleiding en de constructieve feedback gedurende het hele traject.

Daarnaast wil ik mijn dank uiten aan Benny Claes en Kevin Valkenborg van het bedrijf Bewel. Hun bereidheid om hun expertise te delen en de deuren van hun organisatie open te stellen, heeft een onmisbare bijdrage geleverd aan de totstandkoming van deze masterproef.

Een woord van dank is ook op zijn plaats voor mijn vrienden van de universiteit. De gezamenlijke studie-uren, de aanmoediging en de vriendschap hebben de uitdagingen van het afstuderen aanzienlijk verlicht.

Tot slot wil ik mijn naaste familie bedanken voor hun voortdurende steun en vertrouwen. Een speciale dank aan mijn ouders en zus, die me altijd hebben aangemoedigd en geholpen wanneer dat nodig was. Zonder hun steun was dit alles niet mogelijk geweest.

Inhoudsopgave

Lijst van tabellen

Lijst van figuren

Abstract

Deze masterproef beschrijft de ontwikkeling van een virtual reality (VR)-gebaseerd systeem om de objectiviteit en schaalbaarheid van competentiebeoordelingen voor medewerkers van Bewel, een organisatie in Limburg die personen met arbeidsproblemen ondersteunt, te verbeteren. De huidige werkwijze steunt op handmatige, observatiegestuurde tests die subjectief worden beoordeeld door begeleiders, wat leidt tot variabele uitkomsten en een hoge tijdsinvestering. Als oplossing van deze beperkingen is in Unity, gebruikmakend van de OpenXR-standaard, een VR-omgeving ontwikkeld om drie screeningopdrachten te simuleren: inhoudsinspectie van een doos, het gebruiken van een zeefdrukmachine en het stapelen van emmers volgens een gegeven patroon. Het systeem registreert gebruikersacties en simuleert interacties via VR-handtracking. Daarnaast werd een begeleidersdashboard ontwikkeld waarmee testresultaten per opdracht gevisualiseerd en geëvalueerd kunnen worden. Dit dashboard ondersteunt objectieve metingen door de gemeten data te visualiseren en maakt het mogelijk om scores automatisch te genereren. Uit de evaluatie met maatwerkers en begeleiders blijkt dat de VR-toepassing zowel technisch als functioneel haalbaar is. De resultaten tonen aan dat het systeem de subjectiviteit in de beoordeling vermindert, doordat het de zekerheid van de beoordelaars verhoogt wat resulteert in een meer objectieve en uniforme evaluatie. De case study bevestigt het potentieel van VR als aanvullend screeningsinstrument binnen maatwerkbedrijven.

Abstract in English

This master's thesis describes the development of a virtual reality (VR)-based system designed to improve the objectivity and scalability of competency assessments for employees at Bewel, an organization in Limburg that supports individuals with employment challenges. The current assessment method relies on manual, observation-based tests that are subjectively evaluated by supervisors, resulting in inconsistent outcomes and a high time investment. To address these limitations, a VR environment was developed in Unity using the OpenXR standard to simulate three screening tasks: inspecting the contents of a box, operating a screen-printing machine, and stacking buckets according to a given pattern. The system tracks user actions and simulates interactions through VR hand tracking. Additionally, a supervisor dashboard was created to visualize and evaluate test results per task. This dashboard supports objective measurements by visualizing the measured data and makes it possible to automatically generate scores. Evaluation with employees and supervisors shows that the VR application is both technically and functionally feasible. The results show that the system reduces subjectivity in the assessment by increasing the certainty of the assessors, which results in a more objective and uniform evaluation. The case study confirms the potential of VR as a supplementary screening tool for sheltered employment organizations.

Hoofdstuk 1

Inleiding

1.1 Situering

Bewel is een organisatie in Limburg die werkgelegenheid biedt aan mensen met een afstand tot de arbeidsmarkt, zoals mensen met een handicap of andere werkbelemmeringen. Het bedrijf zet zich in voor de persoonlijke en professionele groei van zijn werknemers. Dit wordt bereikt door nauwe begeleiding en aangepaste trainingsprogramma's die inspelen op de specifieke behoeften van elke werknemer. Een centraal onderdeel van deze aanpak is het opleidings- en beoordelingskader dat wordt gebruikt om de ontwikkeling van de competenties van werknemers te meten.

Op dit moment evalueert Bewel technische vaardigheden aan de hand van praktische, praktijkgerichte tests: begeleiders observeren werknemers bij het uitvoeren van taken en kennen scores toe op een schaal van 1 tot 4. Hoewel dit proces directe menselijke feedback geeft, is het sterk afhankelijk van de tijd en het subjectieve oordeel van individuele begeleiders. Hierdoor is het evaluatieproces arbeidsintensief en kunnen de resultaten per beoordelaar verschillen.

Technologische vooruitgang biedt nieuwe mogelijkheden om dergelijke trainingssystemen te verbeteren. Met name immersieve tools zoals virtual reality (VR) en augmented reality (AR) kunnen realistische werkomgevingen simuleren en prestatiegegevens op een gestandaardiseerde manier vastleggen. In deze thesis wordt onderzocht hoe een op VR gebaseerd evaluatiesysteem de beperkingen van de huidige aanpak zou kunnen aanpakken en de missie van Bewel zou kunnen ondersteunen. Er wordt onderzocht of deze technologieën gebruikt kunnen worden om de beoordeling van competenties objectiever, schaalbaarder en verrijkender te maken voor zowel werknemers als begeleiders.

1.2 Probleemstelling

Het huidige evaluatieproces bij Bewel is waardevol, maar kent een aantal belangrijke uitdagingen. De afhankelijkheid van menselijke begeleiders zorgt voor variabiliteit in de beoordelingen: subjectieve beoordelingen, menselijke fouten en verschillen in beoordelingsstijl kunnen leiden tot inconsistente scores. Daarnaast is het proces arbeidsintensief en tijdrovend. Begeleiders moeten veel tijd besteden aan één-op-één evaluaties, waardoor het aantal beoordelingen dat kan worden uitgevoerd beperkt is en er minder tijd beschikbaar is voor aanvullende persoonlijke begeleiding of training.

Een andere belangrijke uitdaging is dat de bestaande methodologie beperkt is tot wat fysiek haalbaar is in de werkplaats. Deze beperking beperkt het aantal vaardigheden en scenario's dat kan worden geoefend en geëvalueerd. Het is bijvoorbeeld niet altijd mogelijk om werknemers te trainen op elk product of proces dat Bewel verwerkt, vanwege ruimte beperkingen of de complexiteit van bepaalde assemblagetaken.

In deze context is de centrale onderzoeksvraag: Hoe kan een VR-gebaseerd screeningssysteem gebruikt worden om de evaluatie van competenties bij de werknemers van Bewel te objectiveren, automatiseren en verrijken? Om deze vraag te beantwoorden, onderzoekt het onderzoek de volgende subvragen:

1. Welke technische én niet-technische competenties kunnen effectief worden gemeten en getraind via VR?
2. Hoe kunnen technologieën zoals eye-tracking en hand-tracking relevante gedragsdata verzamelen?
3. Hoe kan een algoritme op basis van deze data een score toekennen die equivalent of beter is dan de huidige beoordeling door begeleiders?
4. Welke voordelen en beperkingen, zowel technisch als ethisch, brengt de implementatie van een dergelijk systeem met zich mee?
5. Hoe wordt VR-gebaseerde evaluatie gepercipieerd door werknemers? Vinden zij deze methoden waardevol, correct en aangenaam in vergelijking met fysieke testen in combinatie met een menselijke evaluator?

1.3 Doelstellingen

De hoofddoelstelling van dit project is het ontwikkelen van een VR-gebaseerd systeem dat objectieve, schaalbare en herhaalbare evaluaties en trainingen mogelijk maakt voor maatwerkers bij Bewel. Dit systeem moet niet alleen een subset van de huidige evaluatiemethoden ondersteunen, maar ook nieuwe mogelijkheden bieden die met traditionele methoden niet of moeilijk haalbaar zijn. Het is nadrukkelijk niet de intentie om mensen te vervangen, maar juist om hen te ondersteunen bij het maken van een meer objectieve en consistente beoordeling.

Om deze hoofddoelstelling te realiseren, zijn er enkele specifieke doelstellingen:

1. Ontwerpen van trainingsscenario's in VR: Identificeren en digitaliseren van realistische werkprocessen, vaardigheden en andere relevante aspecten zoals veiligheidstraining of attitudevorming, die representatief zijn voor de taken en vereisten binnen Bewel.
2. Integratie van sensortechnologieën: Gebruik maken van eye-tracking, hand-tracking of andere sensoren om realtime data te verzamelen over de acties en prestaties van de gebruiker.
3. Ontwikkeling van een beoordelingsalgoritme: Dit algoritme moet gedragspatronen analyseren en objectieve scores genereren. Deze scores dienen consistent en betrouwbaar te zijn en moeten de huidige evaluatiemethoden aanvullen of overtreffen.
4. Validatie en gebruikersstudie: Het systeem testen met werknemers en begeleiders van Bewel om de bruikbaarheid, betrouwbaarheid en acceptatie te beoordelen.

5. Competentie-evolutie opvolgen: Door het gedigitaliseerde systeem kunnen testen en trainingen periodiek herhaald worden, waardoor de ontwikkeling van vaardigheden en competenties van maatwerkers door de tijd gemonitord kan worden.
6. Praktische en ethische evaluatie: Analyseren van het ethische kader voor implementatie binnen een inclusieve werkomgeving.

1.4 Vooruitblik

Deze masterthesis start in Hoofdstuk 2: Bronnenstudie met het leggen van een stevig theoretisch en methodologisch fundament. Dit hoofdstuk analyseert de beperkingen van traditionele beoordelingsmethoden en bouwt een argument op voor VR als oplossing. Er wordt onderzocht hoe sensortechnologie en machine learning objectieve metingen mogelijk maken en hoe VR een gestandaardiseerde en veilige omgeving biedt om deze technologieën te integreren. Tevens worden cruciale ontwerpprincipes en validatiestrategieën voor effectieve VR-screenings besproken, die als leidraad dienen voor de verdere ontwikkeling.

Vervolgens wordt in Hoofdstuk 3: Implementatie de vertaalslag gemaakt van theorie naar praktijk. Dit hoofdstuk beschrijft gedetailleerd het technische ontwikkelingsproces van het VR-systeem, beginnend bij een grondige analyse van de noden bij Bewel. De architectuur van de software, de keuzes voor de technologische basis en de ontwikkeling van een op maat gemaakt interactiesysteem voor nauwkeurige objectmanipulatie worden toegelicht. Daarnaast wordt de architectuur voor data-acquisitie, het geautomatiseerde beoordelingsmodel en de uiteindelijke realisatie van het resultatendashboard voor begeleiders uiteengezet.

Met het voltooide systeem wordt in Hoofdstuk 4: Evaluatie de methodologie voor de empirische validering beschreven. Dit hoofdstuk legt de opzet van de tweeledige evaluatiestudie uit: enerzijds de toetsing van de VR-applicatie bij de doelgroep van maatwerkers en anderzijds de evaluatie van het resultatendashboard door begeleiders.

De bevindingen van deze evaluatie worden gepresenteerd in Hoofdstuk 5: Resultaten. Dit hoofdstuk analyseert zowel de kwantitatieve als de kwalitatieve data die tijdens de studies zijn verzameld. Er wordt gerapporteerd over de gebruikerservaring van de VR-screening en de impact van het dashboard op de objectiviteit, zekerheid en het oordeel van de beoordelaars.

Ten slotte worden in Hoofdstuk 6: Besluit alle onderzoekslijnen samengebracht. Dit afsluitende hoofdstuk interpreteert de resultaten in een breder kader, bekijkt de theoretische en praktische implicaties van het onderzoek, erkent de beperkingen van de studie en formuleert concrete aanbevelingen voor zowel de doorontwikkeling van het systeem als voor toekomstig onderzoek in dit domein.

Hoofdstuk 2

Bronnenstudie

In dit hoofdstuk wordt de evolutie van beoordelingsmethoden geanalyseerd: van subjectieve observatie naar kwantitatieve metingen met technologische ondersteuning. Deze analyse leidt tot de conclusie dat VR een uitstekende oplossing biedt voor dit vraagstuk. VR heeft unieke voordelen, zoals de mogelijkheid om een realistische werkomgeving na te bootsen, de testomgeving volledig te standaardiseren en het proces van dataverzameling te automatiseren. Hierdoor kan VR een nauwkeuriger en volledigere beeld geven van de competenties die van belang zijn voor tewerkstelling in een maatwerkbedrijf.

2.1 De beperkingen van conventionele screening

De traditionele methoden voor het beoordelen van vaardigheden kunnen worden onderverdeeld in twee categorieën: productgebaseerde en procesgebaseerde metingen [1]. Productgebaseerde metingen richten zich op de kwantitatieve uitkomst van een handeling, zoals de afstand van een worp of de tijd die nodig is om een taak te voltooien. Hoewel deze metingen eenvoudig te objectiveren zijn, bieden ze geen inzicht in de kwaliteit van de beweging. Procesgebaseerde metingen daarentegen focussen op de kwalitatieve aspecten van de uitvoering, zoals de techniek van een beweging [2]. Denk hierbij aan een teamleider die met een checklist de handelingen van een operator aan de assemblagelijn observeert, of een expert die de laskwaliteit van een nieuwe medewerker beoordeelt op basis van diens techniek. Dit is waar de meest significante beperkingen van conventionele methoden aan het licht komen.

De kern van het probleem is de afhankelijkheid van een menselijke observator, wat onvermijdelijk leidt tot subjectiviteit. Zelfs ervaren beoordelaars hebben moeite om bekwaamheidscriteria consistent en accuraat te identificeren [1], [3], wat resulteert in een lage betrouwbaarheid. Deze betrouwbaarheid neemt verder af wanneer men van het algemene vaardigheidsniveau naar de specifieke criteria binnen een vaardigheid kijkt. Traditionele veiligheidstrainingen, zoals lezingen en video's, illustreren een gerelateerd probleem: ze creëren een passieve leeromgeving [4] en bieden vaak geen gestructureerde feedback [5], [6], wat een objectieve meting van competentie bemoeilijkt.

Daarnaast zijn deze methoden arbeidsintensief. Het live scoren van prestaties is een tijdrovend en kostbaar proces. Deze hoge personeelsinzet en de benodigde training voor beoordelaars belemmeren de haalbaarheid en schaalbaarheid van grootschalige screenings [1], [3], [7], [8]. Het

(gedeeltelijk) automatiseren van de beoordeling zou examinatoren echter in staat stellen hun cognitieve middelen effectiever in te zetten, waardoor ze zich kunnen richten op vaardigheden van een hogere orde, zoals communicatie [8].

Tot slot is er een gebrek aan context. Traditionele testen vinden vaak plaats in een gecontroleerde, maar artificiële setting die de complexiteit en dynamische factoren van een echte werkomgeving mist. Het is bijvoorbeeld niet efficiënt om een dure of cruciale machine uit de productielijn te halen voor screeningsdoeleinden. Dit zou niet alleen de productie stilleggen [5], maar brengt ook veiligheidsrisico's [9] en aanzienlijke kosten met zich mee. Het resultaat is een test die vaardigheden in isolatie meet [4], losgekoppeld van de omgeving waarin ze moeten worden toegepast. Hierdoor wordt de voorspellende waarde van de test aanzienlijk beperkt.

2.2 De rol van sensortechnologie en machine learning

De beperkingen van traditionele observatie hebben geleid tot een zoektocht naar objectieve, kwantitatieve meetmethoden. De introductie van sensortechnologie, zoals Inertial Measurement Units (IMU's) en accelerometers, markeert hierin een paradigmaverschuiving. Deze sensoren, die typisch accelerometers en gyroscopen bevatten, maken het mogelijk om objectieve data over menselijke beweging te verzamelen [2], zoals driedimensionale versnellingen en hoeksnelheden [1], [7], [10]. Deze technologie verschuift de focus van een subjectief oordeel naar een data-gedreven analyse van de beweging zelf [3].

De kracht van deze benadering komt tot uiting wanneer deze verwerkte data worden geanalyseerd met Machine Learning (ML) algoritmes. Modellen zoals Hidden Markov Models (HMM), Support Vector Machines (SVM) en Artificial Neural Networks (ANN) [3] worden getraind op datasets waarin de prestaties gelabeld zijn (bv. 'beginner' versus 'expert') [10], [3]. Deze algoritmes leren vervolgens om de subtiele, vaak voor het menselijk oog onzichtbare, patronen in de kinematische data te herkennen die geassocieerd zijn met verschillende vaardigheidsniveaus. Studies tonen aan dat dergelijke systemen met een hoge nauwkeurigheid, vaak boven de 80% of zelfs 90% [1], [10], [3], onderscheid kunnen maken tussen vaardigheidsniveaus. Deze methode elimineert niet alleen de subjectiviteit van de menselijke beoordelaar, maar reduceert ook drastisch de tijd en kosten die gepaard gaan met analyse [1], [3].

2.3 VR als screening instrument

VR onderscheidt zich van andere technologieën door een unieke combinatie van eigenschappen die het bijzonder geschikt maken voor screening doeleinden. De meest fundamentele eigenschap is immersie [4], [11], het vermogen van de technologie om de gebruiker volledig onder te dompelen in een computer-gegenereerde 3D-omgeving door de fysieke wereld visueel en auditief af te sluiten. Deze immersie creëert een sterk gevoel van aanwezigheid [12], het subjectieve gevoel daadwerkelijk aanwezig te zijn in de virtuele wereld. Dit realisme is belangrijk, omdat het authentieke reacties en gedragingen kan uitlokken die sterk lijken op die in de echte wereld [13], [14], wat essentieel is voor een valide assessment.

Een tweede voordeel is de controleerbaarheid en standaardisatie van de omgeving. In tegenstelling tot de onvoorspelbare en variabele omstandigheden van een echte werkplek, biedt VR een perfect gestandaardiseerde testomgeving. Elke kandidaat voert de taak uit onder exact dezelfde

condities, met dezelfde instructies, materialen en mogelijke afleidingen. Dit maakt het mogelijk om situaties die in het echte leven moeilijk of gevaarlijk zijn om na te bootsen, zoals noodscenario's, veilig en herhaaldelijk te repliceren zonder risico's voor de deelnemers [8]. Dit garandeert de vergelijkbaarheid en eerlijkheid van de screening. Bovendien kan de moeilijkheidsgraad van de taken dynamisch en in real-time worden aangepast aan de prestaties van de gebruiker, wat gepersonaliseerde en adaptieve screening mogelijk maakt.

Een van de voordelen van VR in de context van assessment is de hoge ecologische validiteit [4], [12]: de mate waarin de prestaties en bevindingen binnen de testomgeving generaliseerbaar zijn naar alledaagse, reële situaties. Traditionele tests, zoals pen-en-papiertaken of abstractere testen, hebben vaak een lage ecologische validiteit [4]. Ze meten cognitieve of motorische functies in isolatie, los van de complexe en dynamische context waarin deze vaardigheden in de praktijk moeten worden toegepast.

VR-screenings overkomen deze beperking door deelnemers onder te dompelen in realistische, interactieve scenario's die de complexiteit van alledaagse taken simuleren. In plaats van een abstracte test voor hand-oogcoördinatie, kan een VR-screening een deelnemer vragen een specifieke assemblagetaak uit te voeren die identiek is aan een taak op de werkvloer. Een dergelijke taakspecifieke simulatie spreekt niet alleen de motorische vaardigheid aan, maar integreert ook andere relevante competenties. Hierdoor is de voorspellende waarde voor de uiteindelijke werkprestatie significant hoger dan die van een generieke, contextloze test.

De voordelen van VR strekken zich bovendien uit tot training. Studies tonen aan dat VR-training minstens zo efficiënt of zelfs effectiever is dan traditionele methoden zoals video's en lezingen, zowel voor de initiële kennisverwerving als voor het behoud van kennis op de lange termijn [5], [6], [8].

2.4 Het kwantificeren van prestaties in VR

VR-screening maakt het mogelijk om prestaties automatisch en objectief te kwantificeren door middel van datapunten die door het VR-systeem in real-time en zonder menselijke tussenkomst worden gelogd [4], [14]. Dit proces garandeert objectiviteit, perfecte reproduceerbaarheid en biedt de mogelijkheid voor gedetailleerde, onmiddellijke feedback [15].

Een eerste voorbeeld van een datapunt zijn kinematische metingen [7], [10]. Deze meten de kwaliteit en efficiëntie van de beweging. Voorbeelden zijn de totale afgelegde afstand van de handen of controllers, een kortere afstand duidt op een efficiëntere beweging [15], de gemiddelde en maximale snelheid en versnelling en de versnellingsverandering [7], wat de maat is voor de vloeiendheid van de beweging. Een lage versnellingsverandering wijst op soepele, beheerste bewegingen, terwijl een hoge versnellingsverandering duidt op schokkerige, aarzelende acties.

Een tweede voorbeeld is tijdmetingen [5], [10]. De meest voor de hand liggende meting is de totale taakvoltooiingstijd. Echter, meer gedetailleerde analyses zijn mogelijk door de tijd per subtaak te meten [11], of de tijd dat gereedschap of hand in contact is met een specifiek object. Dit kan inefficiënties in specifieke fases van een taak blootleggen.

Ten slotte zijn er taakspecifieke metingen [10]. Deze meten de nauwkeurigheid en het eindresultaat van de taak [5]. Dit zijn de meest directe indicatoren van succes. Voorbeelden zijn het

totale aantal gemaakte fouten, de precisie van plaatsing, volgorde waarin handelingen worden uitgevoerd of het aantal keren dat de gebruiker om hulp of instructies vraagt[9].

2.4.1 Meten tegen een expertmodel

Een directe methode voor beoordeling is het vergelijken van een prestatie met een vooraf gedefinieerd 'ideaal' model [14]. De techniek Dynamic Time Warping (DTW) is hier bijzonder geschikt voor [1], [3]. DTW meet de gelijkenis tussen twee temporele reeksen die in snelheid kunnen variëren. In de studie naar motorische vaardigheden bij kinderen wordt DTW gebruikt om de Euclidische afstand te meten tussen het acceleratiesignaal van een kind en een 'goed' signaalmodel (een 'centroid' berekend uit de prestaties van vaardige kinderen). Dit maakt een robuuste vergelijking mogelijk, zelfs als het ene kind de beweging sneller of langzamer uitvoert dan het andere.

2.4.2 Machine learning

Een van de manieren om het beoordelingsproces te automatiseren is door het gebruik van supervised learning [10]. Hierbij wordt een model getraind op een dataset die vooraf is gelabeld door menselijke experts (bv. 'beginner', 'gevorderd', 'expert'). Het model leert de relatie tussen de input en het correcte vaardigheidslabel [3].

- Support Vector Machines (SVM): Een krachtig en veelgebruikt algoritme dat een optimaal scheidingsvlak (hyperplane) zoekt dat de verschillende vaardigheidsklassen van elkaar scheidt. SVM's zijn effectief, zelfs met de relatief kleine datasets die vaak in medische trainingstudies voorkomen.
- Artificial Neural Networks (ANN): Deze modellen, geïnspireerd op de werking van het menselijk brein, kunnen zeer complexe, niet-lineaire relaties in de data leren. Ze winnen snel aan populariteit en behalen vaak state-of-the-art nauwkeurigheid, met gerapporteerde scores van meer dan 80-90% [1], [10], [3] bij het voorspellen van trainingsniveaus in chirurgische simulaties.
- Hidden Markov Models (HMM): HMM's zijn probabilistische modellen die uitermate geschikt zijn voor het analyseren van sequentiële data. Ze kunnen de onderliggende structuur van een taak ('verborgen' staten zoals 'instrument benaderen', 'hechten', 'knippen') afleiden uit de observeerbare sensordata (de kinematische metingen). Door de volgorde, duur en overgangen tussen deze staten te analyseren, kan een HMM de efficiëntie en de logische 'flow' van de taakuitvoering beoordelen.

2.5 Multimodale data

Hoewel de vorige metingen een objectief beeld geven van wat een gebruiker doet, verklaren ze niet altijd waarom zij op een bepaalde manier presteren. Een trage taakuitvoering kan bijvoorbeeld wijzen op een gebrek aan motorische vaardigheid, maar kan evengoed het gevolg zijn van hoge stress, verwarring door de instructies, of een hoge cognitieve belasting. Om een hier inzicht in te krijgen, kunnen deze metingen worden verrijkt met data uit andere, multimodale bronnen.

Door de oogbewegingen van de gebruiker te volgen, kan worden vastgesteld waar de visuele

aandacht op gericht is [4]. Leest de gebruiker de instructies? Wordt de aandacht afgeleid door irrelevante stimuli? Kijkt de gebruiker vooruit naar de volgende stap in het assemblageproces?

sommige sensoren meten direct de fysiologische toestand van de gebruiker. Galvanic Skin Response (GSR) [4] is een indicator voor fysiologische arousal en stress. Elektro-encefalografie (EEG) [4] kan de hersenactiviteit meten en zo inzicht geven in de cognitieve belasting (cognitive load) en mentale vermoeidheid.

Het combineren van al deze databronnen leidt tot een veel genuanceerder competentieprofiel [12]. Stel dat een gebruiker veel fouten maakt en een inefficiënt bewegingspad volgt. Zonder extra data zou de conclusie kunnen zijn dat de motorische vaardigheden van de persoon tekortschieten. Als de EEG-data echter tegelijkertijd een hoge cognitieve belasting aantonen, verschuift de interpretatie. Het probleem is dan mogelijk niet primair motorisch, maar cognitief: de instructies zijn te complex of de taak vereist te veel gelijktijdige mentale operaties. De aanbeveling is dan niet simpelweg 'meer motorische oefening', maar eerder 'vereenvoudiging van de instructies' of 'het opdelen van de taak in kleinere stappen'. Deze multimodale aanpak transformeert de screening van een loutere beoordeling naar een krachtig diagnostisch instrument.

2.6 Ontwerpprincipes

Het ontwerpen van een effectieve VR-screening begint bij een diepgaand begrip van hoe mensen nieuwe motorische en procedurele vaardigheden leren. De unieke eigenschappen van VR maken het mogelijk om verschillende leerstrategieën te implementeren en te combineren. Deze sectie legt de theoretische basis door de meest relevante leermethoden te analyseren en de cruciale rol van mentale modellen en vaardigheidstransfer te belichten.

2.6.1 Actief versus observationeel leren

Voor het aanleren van manuele vaardigheden in VR kunnen twee fundamentele benaderingen worden onderscheiden: actief leren en observationeel leren. Hoewel ze verschillend zijn in uitvoering, blijken ze complementaire voordelen te bieden.

Actief Leren is de meest gangbare en traditionele aanpak in industriële VR-trainingen. Bij deze methode leert de gebruiker door de taak zelf uit te voeren vanuit een eerstepersoonsperspectief. De gebruiker interageert direct met de virtuele objecten en gereedschappen, vaak ondersteund door multimodale begeleiding zoals tekstinstructies, oplichtende objecten of geanimeerde aanwijzingen. Deze 'hands-on' benadering is onmisbaar voor het ontwikkelen van motorische vaardigheden en het opbouwen van zogenaamd 'muscle memory'. Het is met name essentieel in de latere fasen van het leerproces, waarin de vaardigheid door herhaling wordt verfijnd en geautomatiseerd.

Observationeel Leren is een alternatieve methode waarbij de gebruiker leert door een expert in VR, vaak gerepresenteerd als een avatar, de taak correct te zien uitvoeren. Bij de inzet van een demonstrator-avatar rijst de vraag hoe deze eruit moet zien. Onderzoek naar de impact van visuele gelijkenis tussen de avatar en de gebruiker toonde echter geen significant leervoordeel aan voor avatars die sterker op de gebruiker leken. Realistische zelf-avatars kunnen zelfs afleidend werken door de nieuwigheid ervan [11] of door het 'uncanny valley'-effect [14], waarbij een bijnamenselijke representatie als ongemakkelijk wordt ervaren. Deze aanpak maakt gebruik van het aangeboren menselijke vermogen om te leren door imitatie en is significant minder cognitief

belastend dan actief leren, vooral voor beginners [11]. Onderzoek toont aan dat het integreren van een voorafgaande observatiefase de totale trainingstijd aanzienlijk kan verkorten, zonder dat dit ten koste gaat van de uiteindelijke prestatie op de reële taak [16]. In plaats van direct te moeten handelen, kan de gebruiker zijn volledige aandacht richten op het begrijpen van de procedure en de onderliggende strategie.

De effectiviteit van observationeel leren is direct gekoppeld aan het verminderen van de initiële cognitieve belasting [16]. Een beginnende gebruiker die direct een complexe assemblagetaak moet uitvoeren (actief leren), wordt geconfronteerd met een driedubbele taak: het begrijpen van de procedure, het manipuleren van de objecten en het bedienen van de VR-interface. Dit kan leiden tot een cognitieve overbelasting die het leerproces hindert. Door de training te starten met een observatiefase, wordt de cognitieve last van de uitvoering tijdelijk weggenomen. Hierdoor kan de gebruiker zijn mentale middelen volledig inzetten voor het vormen van een accuraat mentaal model van de taak. Dit robuuste mentale model vormt de basis voor betere prestaties, met name in complexere en onvoorspelbare situaties. er is bewijs dat observationeel leren, in vergelijking met actief leren, het vermogen van gebruikers significant verbetert om het geleerde over te dragen naar reële taken, zelfs buiten de specifieke context die in VR is aangeleerd [11]. Een gecombineerde aanpak is daarom het meest effectief [11]: de training begint met een observationele fase om een solide cognitieve basis te leggen, gevolgd door actieve, 'hands-on' oefening om de vaardigheid te verfijnen en te internaliseren.

Hoewel de tweedeling tussen actief en observationeel leren primair is ontwikkeld voor training, biedt dit raamwerk ook een waardevolle lens voor het ontwerpen van effectievere screeningtools. Een traditionele screening zou zich beperken tot een 'actieve' aanpak, waarbij de prestatie van een kandidaat op een onbekende taak direct wordt gemeten. Dit meet weliswaar de onmiddellijke vaardigheid, maar kan negatief beïnvloed worden door de cognitieve belasting van het tegelijkertijd moeten begrijpen van de taak én het bedienen van de VR-interface. Door een observationele fase aan de screening vooraf te laten gaan, kan de aard van de meting worden verfijnd. In een dergelijk 'observeer-dan-doe' protocol wordt de kandidaat eerst in staat gesteld een accuraat mentaal model van de taak te vormen door een expert te observeren. De daaropvolgende actieve uitvoering meet dan niet louter de basisvaardigheid, maar veeleer het leervermogen en de efficiëntie waarmee de kandidaat observatie kan omzetten in correcte handelingen. Deze gecombineerde aanpak zorgt voor een zuiverdere meting van het potentieel, omdat de initiële cognitieve last wordt verlaagd en de invloed van onwennigheid met de technologie wordt geminimaliseerd. Zo kan een screening niet alleen meten wat een kandidaat al kan, maar ook hoe snel zij een nieuwe vaardigheid kunnen aanleren.

2.6.2 Skill transfer

Het uiteindelijke doel van de training is niet alleen het kunnen uitvoeren van een reeks handelingen, maar het ontwikkelen van een dieper begrip dat flexibel kan worden toegepast. Hierin spelen het mentale model en het concept van 'skill transfer' centrale rollen.

De ware maatstaf voor het succes van een training is de mate waarin de geleerde vaardigheden kunnen worden toegepast in de praktijk. Dit wordt 'skill transfer' genoemd [12], en er wordt een belangrijk onderscheid gemaakt tussen 'near transfer' en 'far transfer'.

Near Transfer verwijst naar het toepassen van een geleerde vaardigheid in een context die identiek

is aan de trainingsomgeving. Dit is de meest geteste vorm van transfer in VR-onderzoek en is relevant voor sterk gestandaardiseerde, 'gesloten' taken. Far Transfer verwijst naar het vermogen om een geleerde vaardigheid toe te passen en aan te passen in een nieuwe, afwijkende context. Dit kan betekenen dat de onderdelen in een andere volgorde liggen, kleurcodes ontbreken of er omgevingsafleidingen zijn. Far transfer is een veel betere indicator voor prestaties in de echte, dynamische en variabele wereld van een industriële werkvloer [5].

2.6.3 Definiëren van de simulatiegetrouwheid

Bij het ontwerpen van een VR-simulator is de mate van realisme, of fidelity, een cruciale overweging. Hierbij wordt een onderscheid gemaakt tussen fysieke en cognitieve fidelity, die elk een andere rol spelen in het proces.

Fysieke fidelity verwijst naar de mate waarin de simulator de reële wereld visueel, auditief en haptisch nabootst. Een hoge fysieke fidelity, met realistische modellen van gereedschappen en onderdelen, is essentieel voor de herkenbaarheid van de taak en de initiële acceptatie door de gebruiker. Cognitieve fidelity betreft de mate waarin de simulator de cognitieve processen activeert die ook in de reële taak vereist zijn, zoals procedureel redeneren, probleemoplossing en ruimtelijk inzicht.

Analyse toont aan dat een eenzijdige focus op een van beide vormen van fidelity suboptimaal is [16]. Een applicatie met een hoge fysieke fidelity maar lage cognitieve eisen (bv. een systeem waarbij het volgende te gebruiken onderdeel oplicht) meet mogelijk enkel het vermogen van een kandidaat om simpele visuele cues te volgen. Dit resulteert in oppervlakkig leren en heeft een lage voorspellende waarde voor prestaties in de complexe realiteit. Omgekeerd kan een te lage fysieke fidelity de overdracht van de geleerde vaardigheden naar de fysieke, motorische aspecten van de taak belemmeren.

De screening moet meten of een kandidaat de logica van een taak begrijpt en kan toepassen, niet enkel of zij een reeks voorgeprogrammeerde stappen kunnen imiteren. Een hogere cognitieve fidelity, waarbij de gebruiker bijvoorbeeld een 2D-schema moet interpreteren om een 3D-handeling uit te voeren [16], test niet alleen de motoriek, maar ook het ruimtelijk inzicht en het procedureel redeneren. Dit zijn cruciale componenten van manuele vaardigheid in een complexe maakcontext. De voorspellende waarde van de screening voor prestaties in onvoorspelbare, reële situaties wordt hierdoor aanzienlijk verhoogd.

De VR-omgeving moet een hoge mate van fysieke herkenbaarheid bieden (realistische onderdelen, gereedschappen), maar de instructies en feedback moeten zo worden ontworpen dat ze de gebruiker dwingen tot een cognitieve vertaalslag [4]. Een praktisch voorbeeld is het tonen van een assemblage-instructie op een virtueel scherm in de werkruimte, in plaats van het direct manipuleren van de scène met visuele hints [16]. Overmatige begeleiding tijdens een screening kan de validiteit ervan ondermijnen. Constante visuele of haptische cues, zoals oplichtende onderdelen of andere technieken om de hand van de gebruiker te sturen, kunnen ertoe leiden dat de kandidaat de taak succesvol voltooit zonder deze daadwerkelijk te begrijpen [16]. Dit meet dan eerder het vermogen om instructies te volgen dan de onderliggende vaardigheid.

Tijdens de actieve screening dus geen proactieve hulp. Indien de gebruiker vastloopt, kunnen zij via een virtuele knop of een spraakcommando een hint opvragen. Deze hint kan bestaan uit het kort oplichten van het volgende onderdeel of een korte animatie van de volgende stap. Het

aantal keren dat de hulpfunctie wordt gebruikt, is een cruciale meting die inzicht geeft in het probleemoplossend vermogen van de gebruiker.

2.6.4 Ontwerpprincipes voor een effectieve en inclusieve VR-omgeving

Deze sectie beschrijft hoe de virtuele omgeving en de interacties moeten worden vormgegeven om een effectieve, gestructureerde en ondersteunende leerervaring te creëren die is afgestemd op de noden van de doelgroep.

Een VR-applicatie is pas effectief als deze toegankelijk, motiverend en veilig is voor de beoogde gebruikers. De volgende richtlijnen zijn gebaseerd op een systematische review van VR-interventies voor mensen met een intellectuele of ontwikkelingsstoornis en zijn onderverdeeld in de categorieën haalbaarheid, bruikbaarheid en veiligheid.

De haalbaarheid wordt in de eerste plaats bepaald door het wegnemen van praktische en sensorische barrières. Cognitieve vermoeidheid is een belangrijke factor, waardoor het essentieel is om sessies kort en gefocust te houden, met een aanbevolen duur tussen de 15 en 45 minuten [13]. Langere sessies kunnen leiden tot fysieke en mentale vermoeidheid, wat de prestaties kan beïnvloeden. Een grotere uitdaging is de sensorische gevoeligheid van veel gebruikers. Het dragen van een VR-headset kan overweldigend zijn en leiden tot weigering. Om dit te vermijden, is een desensitisatieprotocol cruciaal, waarbij de gebruiker onder professionele begeleiding stapsgewijs aan de technologie wordt blootgesteld. De aanwezigheid van een getrainde begeleider is hierbij onmisbaar om de gebruiker gerust te stellen en het proces in goede banen te leiden.

Zodra de gebruiker vertrouwd is met de hardware, komt de bruikbaarheid van de software naar voren. De interface en bediening moeten zo intuïtief mogelijk zijn om te voorkomen dat de screening de vaardigheid in het omgaan met de technologie meet in plaats van de beoogde manuele vaardigheid. Herhaald gebruik verbetert de vaardigheid met de interface [14], wat het belang van een goede familiarisatiefase onderstreept. Dit kan bereikt worden door het aantal benodigde knoppen te minimaliseren en maximaal gebruik te maken van natuurlijke handbewegingen voor interacties zoals grijpen en plaatsen. De begeleider speelt hierbij ook een cruciale rol in het personaliseren van de ervaring. Dit omvat het aanpassen van de headset voor comfort, het geven van extra verbale uitleg en het inspelen op de specifieke behoeften van de gebruiker.

Een absolute voorwaarde is het waarborgen van de veiligheid, zowel fysiek als psychologisch. Misselijkheid is een veelvoorkomend probleem in VR [4], [12], [8], veroorzaakt door een verschil tussen wat de ogen zien en wat het evenwichtsorgaan voelt. Symptomen zijn onder meer misselijkheid, duizeligheid en hoofdpijn. Dit kan geminimaliseerd worden door applicaties primair voor zittend gebruik te ontwerpen, de gebruiker actieve controle over bewegingen te geven, het door het systeem gestuurde camerabewegingen of verplaatsingen te vermijden en de sessieduur te beperken met regelmatige pauzes [13]. Omdat de headset het zicht op de fysieke omgeving blokkeert, moet de training altijd plaatsvinden in een obstakelvrije ruimte onder actieve supervisie om vallen of botsen te voorkomen. Deze supervisie is tevens essentieel voor de psychologische veiligheid, om de gebruiker te observeren op signalen van ongemak zoals duizeligheid of desoriëntatie. De rol van de begeleider is dus niet alleen ondersteunend bij de start, maar vormt een constante factor in het garanderen van een veilige en effectieve VR-ervaring.

2.7 Structurering van de screening

Een goed gestructureerde sessie is essentieel voor een betrouwbare en gebruiksvriendelijke screening. Een vierfasenmodel, gebaseerd op de methodologie van succesvolle VR-trainingsstudies, wordt aanbevolen.

Fase 1: Familiarisatie: De gebruiker leert de basisbesturing van de VR-applicatie in een neutrale, taak-onafhankelijke omgeving. Dit omvat het oppakken, draaien en plaatsen van objecten. Deze fase verlaagt de cognitieve belasting tijdens de daadwerkelijke taak en minimaliseert de invloed van eerdere VR-ervaring op de prestaties [8], [11].

Fase 2: Demonstratie/Observatie: De gebruiker observeert een neutrale avatar die de volledige assemblagetaak correct en efficiënt uitvoert, zonder de mogelijkheid tot interactie. Dit bouwt het noodzakelijke mentale model op.

Fase 3: Oefening: De gebruiker krijgt de kans om de taak één keer te oefenen. In deze fase kan feedback (bv. objecten lichten groen op bij correcte plaatsing) en de 'hulp op aanvraag'-functie beschikbaar zijn. Dit vormt de brug tussen observatie en de finale beoordeling.

Fase 4: Screening: De gebruiker voert de taak uit zonder enige vorm van hulp of feedback. Dit is de kern van de screening, waarin de prestatiedata worden verzameld.

2.8 Validatie van een VR-screeningtool

Een technisch geavanceerd beoordelingssysteem, zoals een VR-screeningtool, is pas waardevol als het als wetenschappelijk meetinstrument voldoet aan strikte standaarden van betrouwbaarheid en validiteit. Zonder deze garantie is het onmogelijk te garanderen dat de tool op een consistente wijze meet wat het beoogt te meten. Dit vereist een uitgebreid validatieproces dat verder gaat dan alleen technische haalbaarheid.

Allereerst moet de betrouwbaarheid, oftewel de consistentie en reproduceerbaarheid van de meting, worden vastgesteld. Een van de meest directe methoden hiervoor is de test-hertest betrouwbaarheid [14], die nagaat of een individu vergelijkbare scores behaalt bij herhaalde afnames over tijd. Een hoge correlatie duidt op een stabiele meting. Een studie naar een VR-emotieherkenningstaak toonde bijvoorbeeld een uitstekende betrouwbaarheid met correlaties tussen r=.83 en r=.93 [14] over drie sessies. Daarnaast is de interne consistentie van belang, die onderzoekt of verschillende indicatoren binnen één sessie coherent samenhangen. Bij een virtuele assemblagetaak zou men bijvoorbeeld verwachten dat een snellere uitvoeringstijd correleert met een lager aantal gemaakte fouten.

De validiteit is echter de meest cruciale eigenschap en beantwoordt de vraag of de test daadwerkelijk meet wat hij zou moeten meten. Dit begint bij de constructvaliditeit [7], [14], die onderzoekt of de scores van de tool samenhangen met die van bestaande, gevalideerde instrumenten die hetzelfde meten. De resultaten in de literatuur zijn hierbij niet altijd eenduidig. Zo bleek een VR-tool voor emotionele intelligentie significant te correleren met één traditionele vragenlijst, maar niet met een andere [14]. Minstens zo belangrijk is de mate waarin de tool prestaties in de echte wereld kan voorspellen. Een meta-analyse van VR-simulatoren voor robotchirurgie onthulde bijvoorbeeld een significante positieve correlatie (r=0.67) tussen de prestaties op de VR-simulator en de daadwerkelijke prestaties tijdens een operatie [15].

Hier openbaart zich een belangrijke 'validatiekloof' in de literatuur. Veel studies tonen de technische haalbaarheid en de hoge nauwkeurigheid van machine learning-modellen in een gecontroleerde omgeving aan. Een model kan bijvoorbeeld met 95% accuraatheid de labels van een expert reproduceren. Echter, er is een significant gebrek aan externe en longitudinale validatie [5], [10], [8] die deze resultaten verbindt met prestaties in de echte wereld [11]. Het Kirkpatrick-model voor trainingsevaluatie maakt dit tekort inzichtelijk [12]: veel onderzoek blijft steken op niveau 1 (Reactie: vinden gebruikers de tool prettig?) en 2 (Leren: worden ze beter in de VR-taak zelf?), maar evalueert zelden op niveau 3 (Gedrag: wordt de vaardigheid toegepast op het werk?) en 4 (Resultaten: levert dit een positieve impact op voor de organisatie?).

2.9 Vergelijkbare toepassingen

Om de positionering van de in deze masterproef ontwikkelde tool te duiden, is het noodzakelijk deze te situeren binnen het bredere landschap van bestaande VR-toepassingen. Een eerste categorie omvat tools gericht op klinische en cognitieve assessment, die algemene functionele capaciteiten meten in een medische of onderzoekscontext. Een tweede categorie richt zich op de training van interpersoonlijke vaardigheden (soft skills) binnen een beroepsgerichte context. De derde en meest direct vergelijkbare categorie betreft systemen voor de training en assessment van manuele en procedurele vaardigheden in een industriële of beroepsmatige setting.

2.9.1 Klinische screening

Een van de meest prominente en uitvoerig gevalideerde voorbeelden van een VR-gebaseerd assessmentinstrument is de Virtual Reality Functional Capacity Assessment Tool (VRFCAT). Dit instrument is primair ontwikkeld voor toepassing in klinische en farmaceutische onderzoekssettings en heeft als hoofddoel het bieden van een betrouwbare en valide meting van functionele capaciteit. Meer specifiek meet VRFCAT hoe cognitieve beperkingen, bijvoorbeeld geassocieerd met schizofrenie, milde cognitieve stoornis of de ziekte van Parkinson, de uitvoering van instrumentele activiteiten van het dagelijks leven beïnvloeden [17].

Methodologisch is VRFCAT ontworpen om te voldoen aan de strikte eisen van regelgevende instanties zoals de Amerikaanse Food and Drug Administration (FDA). In klinische studies naar medicijnen die de cognitie verbeteren, wordt vaak geëist dat niet enkel een verbetering op een standaard cognitieve test wordt aangetoond, maar ook dat deze cognitieve winst een betekenisvolle impact heeft op vaardigheden die in het dagelijks leven worden gebruikt. VRFCAT dient als een dergelijke maatstaf, door de potentie tot functionele verbetering te meten in een gestandaardiseerde, gesimuleerde omgeving.

Inhoudelijk presenteert de VRFCAT de gebruiker een reeks scenario's die alledaagse taken simuleren, zoals het volgen van een recept in een keuken, het nemen van een bus, boodschappen doen in een supermarkt en de weg terug naar huis vinden. Deze taken worden vaak aangeboden via een niet-immersieve interface, zoals een PC-gebaseerde first-person game of een tablet-applicatie (bekend als VRFCAT-SL). De prestaties worden gekwantificeerd aan de hand van objectieve metingen zoals de totale benodigde tijd en het aantal gemaakte fouten [17].

Een ander voorbeeld is CAVIRE (Cognitive Assessment by VIrtual REality). Dit systeem is ontworpen als een volledig immersief alternatief voor traditionele, pen-en-papier neuropsychologische

tests zoals de Mini-Mental State Examination (MMSE) en de Montreal Cognitive Assessment (MoCA) [18]. Het hoofddoel is de vroege detectie van algemene cognitieve achteruitgang, met name bij ouderen, door prestaties te meten binnen de zes erkende cognitieve domeinen. CAVIRE wordt vaak geassocieerd met CAVE-technologie (Cave Automatic Virtual Environment), een geavanceerde, kamervullende projectie-omgeving die een zeer hoge mate van immersie biedt.

2.9.2 Aanleren en beoordelen van interpersoonlijke vaardigheden

Binnen het domein van technologisch ondersteunde beroepsrevalidatie heeft zich een duidelijke categorie van VR-tools ontwikkeld die zich niet richt op manuele of technische vaardigheden, maar op interpersoonlijke of 'soft skills'. Een voorbeeld hiervan is het platform Bodyswaps.

Bodyswaps is ontworpen om gebruikers een veilige, immersieve omgeving te bieden waarin zij cruciale sociale en communicatieve vaardigheden kunnen oefenen en ontwikkelen. De focus ligt op competenties zoals actieve luistervaardigheid, assertiviteit, effectieve communicatie en het voeren van sollicitatiegesprekken. Na het uitvoeren van een interactie in een scenario, bijvoorbeeld een gesprek met een virtuele collega, krijgt de gebruiker gepersonaliseerde feedback door middel van AI.

2.9.3 Aanleren en beoordelen van manuele vaardigheden

De meest direct vergelijkbare categorie van VR-toepassingen omvat systemen die ontworpen zijn voor het aanleren of beoordelen van manuele en procedurele taken in industriële, beroepsmatige contexten. Veel commerciële en academische systemen in dit domein zijn primair ontworpen als trainingstools. Voorbeelden hiervan worden aangeboden door bedrijven als Strivr, IQ3Connect en Unity Industry. Platformen zoals Strivr verzamelen meer dan 100 datapunten per seconde, waaronder gedetailleerde tracking van hoofdbewegingen, oogbewegingen, interacties en zelfs spraakanalyse. Deze enorme hoeveelheid data wordt verwerkt en gepresenteerd in dashboards die zijn gericht op managers en HR-afdelingen. De output wordt vaak geïntegreerd met bestaande Learning Management Systems, waardoor de voortgang en prestaties van medewerkers op grote schaal kunnen worden gemonitord. Het doel is het optimaliseren van de efficiëntie, het verlagen van trainingskosten en het verbeteren van de veiligheid en prestaties op de werkvloer. De onderliggende algoritmes die de ruwe data omzetten in een prestatiescore of een competentieprofiel zijn vaak propriëtair en niet inzichtelijk of aanpasbaar voor de eindgebruiker. De focus ligt op de uiteindelijke score en niet op de verifieerbaarheid van het proces dat tot dat resultaat heeft geleid.

2.10 Synthese en unieke positionering

De analyse in deze masterproef laat zien dat het ontwikkelde systeem een unieke plaats inneemt binnen het aanbod van VR-screeningtools. In tegenstelling tot klinische toepassingen richt dit systeem zich niet op de meting van cognitieve constructen, maar op het rechtstreeks vastleggen van toegepaste beroepscompetenties. Het verschilt bovendien van andere academische en commerciële systemen doordat het niet bedoeld is als trainingstool, maar specifiek als screeningsinstrument. Verder onderscheidt het zich van commerciële industriële platformen door een transparante 'glass box'-benadering, waarbij de menselijke expert centraal staat en vertrouwen belangrijker is dan ondoorzichtige automatisering.

Het systeem is geen generieke oplossing, maar werd op maat ontwikkeld voor de specifieke evaluatiebehoeften van een Bewel. Het vertaalt bestaande, relevante werkposten naar een digitale en objectieve vorm en biedt zo een antwoord op concrete organisatorische uitdagingen.

De belangrijkste bijdrage van dit onderzoek is de ontwikkeling en validatie van een VR-screeningtool die een duidelijke niche bedient. Het instrument is ontworpen als diagnostisch hulpmiddel voor begeleiders in een maatwerkbedrijf. Het combineert manuele taken die aansluiten bij de werkpraktijk, een interactief dashboard met uitgebreide resultaatweergave en een volledig transparant, door de gebruiker instelbaar scoresysteem. Deze combinatie ondersteunt een data-gedreven dialoog tussen begeleider en medewerker. Zo wordt de subjectiviteit en arbeidsintensiteit van traditionele observatiemethoden verminderd, terwijl de cruciale rol van de menselijke expert bij de interpretatie van de context behouden blijft.

Hoofdstuk 3

Implementatie

Dit hoofdstuk geeft een gedetailleerde technische beschrijving van de ontwikkelde virtual reality (VR) applicatie. Het doel is om de opbouw van de software stapsgewijs uit te leggen, van de technologische basiskeuzes tot de logica van de applicatie.

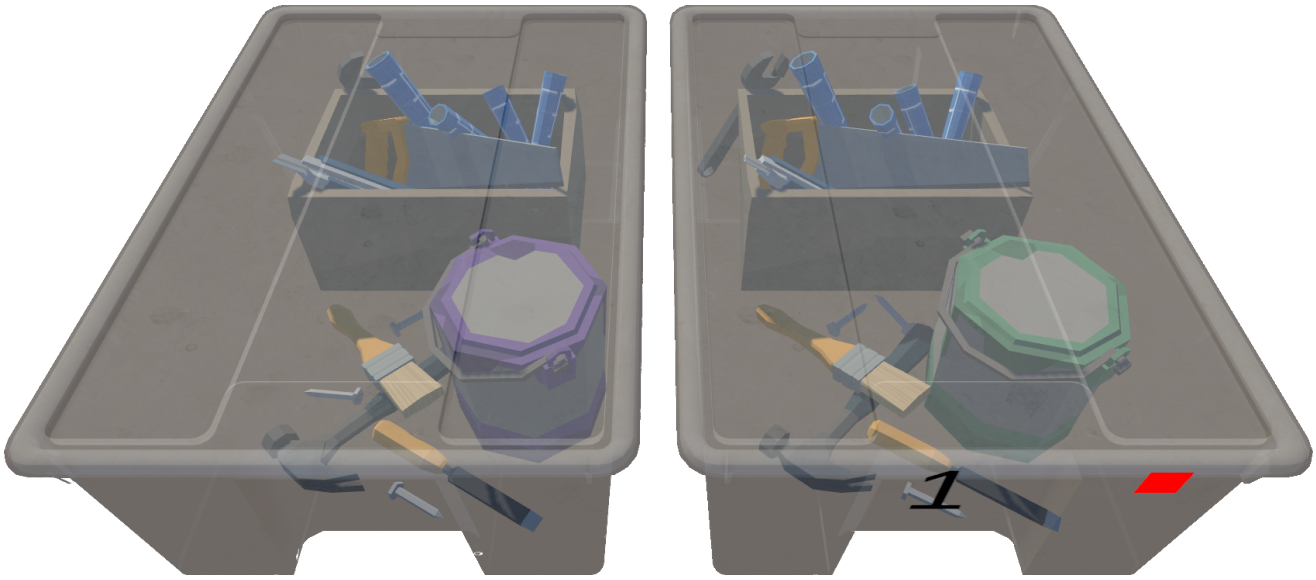
De belangrijkste onderdelen van de applicatie worden een voor een besproken. Eerst worden de gekozen technologieën en ontwerpkeuzes toegelicht die het fundament van het project vormen. Daarna wordt de architectuur van de VR-scenario's geanalyseerd. Een belangrijk onderdeel is het speciaal ontwikkelde interactiemodel, dat zorgt voor een betrouwbare en voorspelbare omgang met objecten; dit wordt uitgebreid behandeld. Vervolgens wordt de methode voor het verzamelen, opslaan en opnieuw afspelen van data beschreven, wat cruciaal is voor een objectieve analyse. Op basis van deze data wordt het beoordelingssysteem uitgewerkt, dat de handelingen van de gebruiker omzet in meetbare scores. Ten slotte wordt de implementatie van het dashboard voor begeleiders besproken, dat dient als de interface om de verzamelde gegevens te interpreteren.

3.1 Analyse van noden bij Bewel

3.1.1 Evaluatie van bestaande proeven voor VR-digitalisering

Om de geïdentificeerde uitdagingen aan te pakken, werd een analyse uitgevoerd om de geschiktheid van de bestaande proeven bij Bewel voor digitalisering in een VR-omgeving te bepalen. De centrale vraag was niet enkel of een taak technisch gerealiseerd kon worden in VR, maar ook of VR een objectievere, schaalbaarder en meer inzichtelijke methode kon bieden om de beoogde competenties te meten. De analyse resulteerde in een categorisering van taken als 'Geschikt', 'Minder geschikt' of 'Niet geschikt', gebaseerd op de aard van de taak en de huidige technologische mogelijkheden van VR.

Taken die als 'Geschikt' werden beoordeeld, zijn die waarbij de kernactiviteiten bestaan uit visuele controle, sorteren, volgen van een stappenplan en het nabouwen van patronen. De reden achter deze keuze is dat dergelijke taken primair een beroep doen op cognitieve vaardigheden die zich uitstekend lenen voor kwantificering in een virtuele omgeving. Bovendien maakt de technologie het mogelijk om een breed scala aan prestatie-indicatoren automatisch en objectief te registreren. Voorbeelden van dergelijke datapunten zijn de totale taakvoltooiingstijd, het aantal en type gemaakte fouten, de precisie van plaatsing, en de mate waarin de correcte volgorde van



Figuur 3.1: Voorbeeld doos (links) en doos met een fout (rechts).

handelingen wordt gerespecteerd.

Aan de andere kant van het spectrum werden taken die een zeer fijne motoriek, complexe sociale interactie of genuanceerde fysieke en haptische feedback vereisen, als 'Minder geschikt' of 'Niet geschikt' voor dit project geëvalueerd.

3.1.2 Motivatie voor de gekozen VR-screeningsscenario's

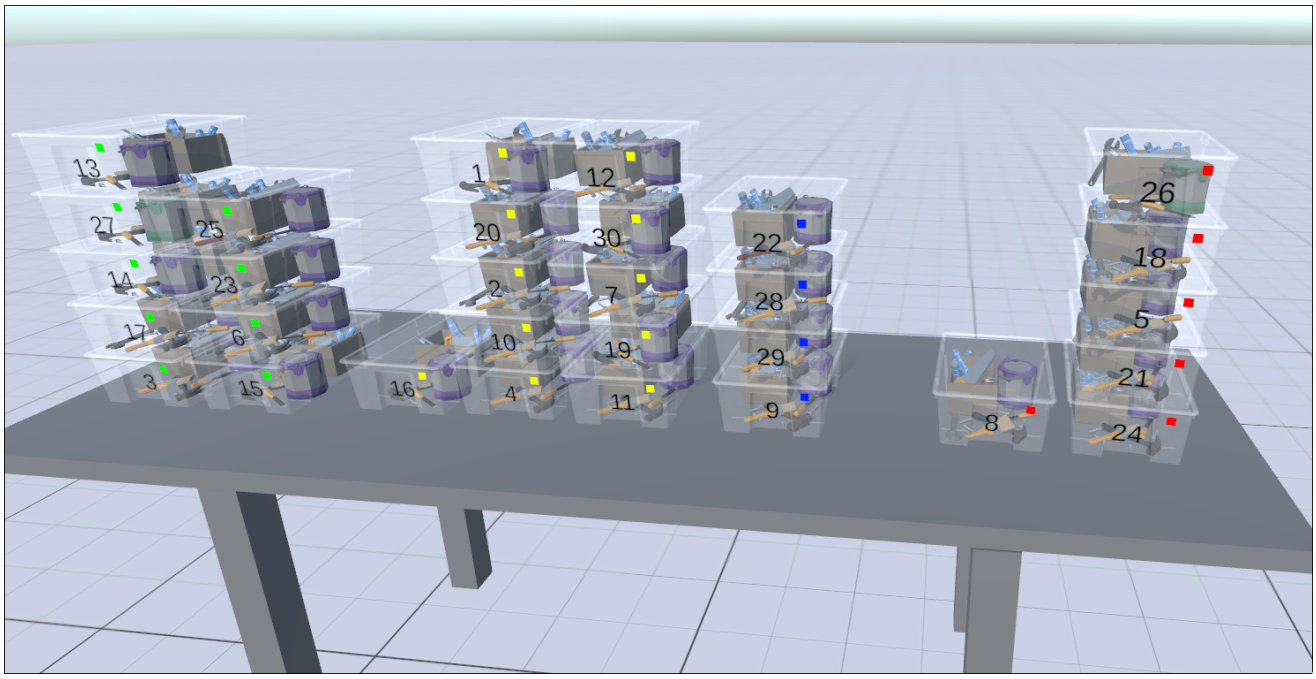
De uiteindelijke selectie van de drie VR-scenario's: de doosinspectie, de zeefdruk en de emmerstapel, is het resultaat van die na overleg met Bewel werd vastgelegd. Het doel van deze selectie was niet enkel om bestaande proeven te digitaliseren, maar om het screening aanbod van Bewel te verrijken en nieuwe mogelijkheden te creëren die met traditionele methoden moeilijk of onmogelijk te realiseren zijn. De drie gekozen scenario's vormen samen een portfolio dat de veelzijdigheid van VR als assessmenttool demonstreert, door respectievelijk een bestaande proef te objectiveren, en een voorheen onmogelijke test mogelijk te maken.

3.1.3 Doosinspectie-screening

In deze sectie wordt het ontwerp en de implementatie van de zogenoemde *doosinspectie-screening* besproken. De screening is gebaseerd op een bestaande proefopstelling binnen Bewel, waarbij werknemers visuele inspecties uitvoeren op verpakkingsdozen en deze correct sorteren.

De screening begint met een korte introductie en instructiefase waarin de gebruiker kennismaakt met de virtuele omgeving, de interactiemogelijkheden en de verwachtingen van de taak. In deze fase krijgt de gebruiker een *voorbeelddoos* gepresenteerd waarin alle componenten correct aanwezig zijn. Deze doos dient als referentie om afwijkingen te detecteren in volgende dozen. De gebruiker kan deze voorbeelddoos te allen tijde opnieuw raadplegen ter ondersteuning van het geheugen en als controlemiddel. Een voorbeeld van deze verschillende dozen zijn te zien op Figuur 3.1.

Na de introductie begint de effectieve screeningsfase. De gebruiker kan door middel van een knop een nieuwe doos opvragen (Figuur 3.3a). Elke doos bevat een aantal objecten die overeenkomen



Figuur 3.2: Voorbeeld van dozen gestapeld op een correcte manier.

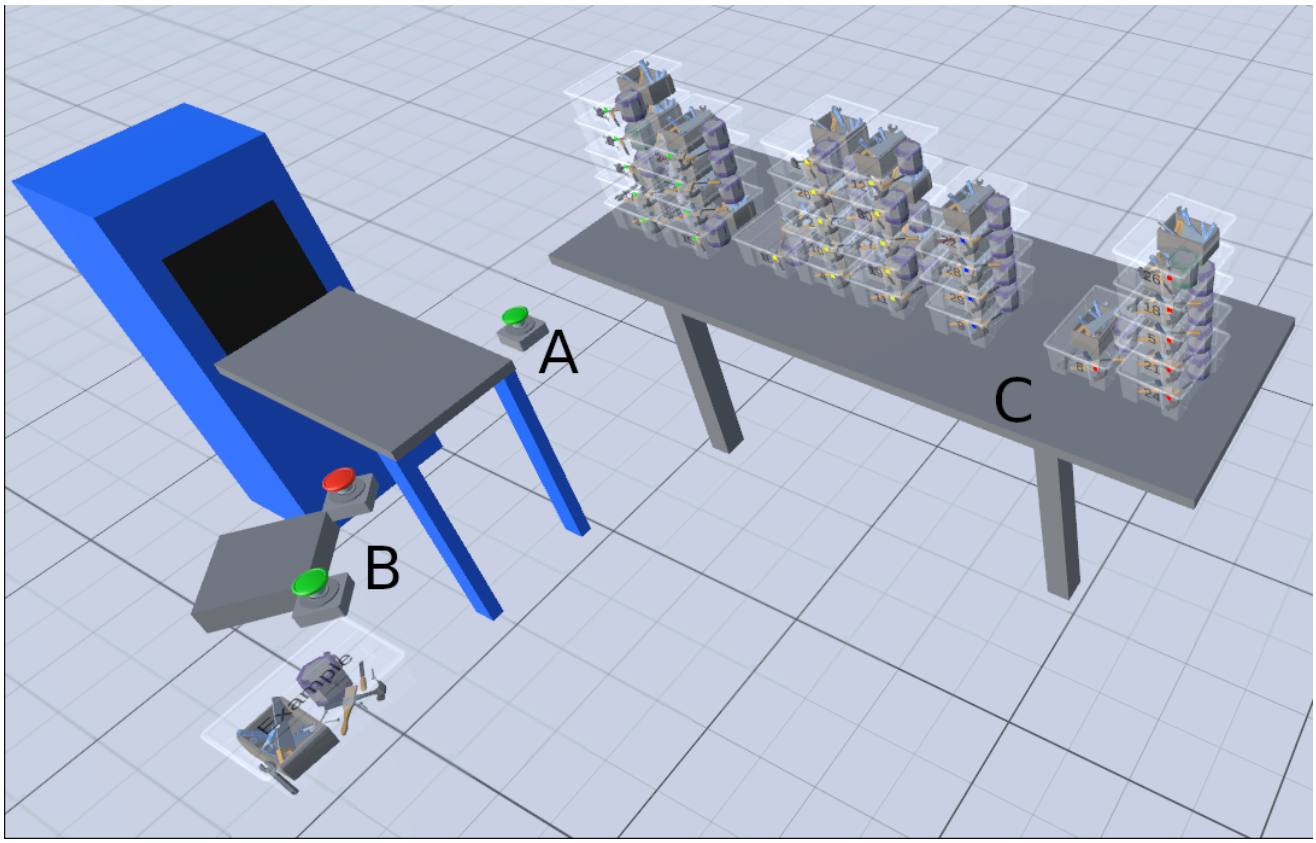
met of afwijken van de voorbeeldinhoud. De gebruiker moet de doos visueel inspecteren en vervolgens een binaire beslissing nemen: "juist" of "fout" (Figuur 3.3b). Indien de doos foutief is, moet de gebruiker dit oordelen op basis van één of meerdere afwijkingen zoals: een overmaat of tekort aan bepaalde objecten, een object dat een afwijkende kleur heeft, of het volledig ontbreken van een verwacht object.

Om de taak realistisch en veelzijdig te maken, is een diversiteit aan fouten ingebouwd. Sommige fouten zijn subtiel, zoals een object dat visueel gelijkt op het correcte item maar net een andere tint heeft. Andere fouten zijn opvallend, zoals een volledig leeg compartiment of een doos met een dubbele hoeveelheid van een item. Deze variabiliteit in moeilijkheidsgraad laat toe om te differentiëren tussen gebruikers op vlak van nauwkeurigheid, snelheid van verwerking, en visuele aandacht.

Na de inspectie- en labelfase, wordt de gebruiker gevraagd om de doos fysiek te verplaatsen en te stapelen op een stapelzone (Figuur 3.3c). Elke doos is aan de buitenzijde voorzien van een kleurmarkering, waarvan vier varianten bestaan: rood, groen, blauw en geel. De gebruiker moet de dozen stapelen in afzonderlijke kolommen, waarbij elke kolom slechts één kleur mag bevatten. Bovendien mag geen enkele stapel hoger worden dan vijf dozen. Figuur 3.2 toont hoe een correcte stapeling eruit kan zien. Deze vereiste introduceert een element van ruimtelijke planning en geheugen, aangezien de gebruiker zich moet herinneren welke stapels reeds gebruikt zijn en hoeveel dozen zich reeds in elke kolom bevinden.

3.1.4 Zeefdruk-screening

Het tweede scenario dat werd geïmplementeerd, is de simulatie van een zeefdrukmachine. Deze taak is gebaseerd op een bestaande werkpost binnen Bewel, waar werknemers verantwoordelijk zijn voor het manueel bedrukken van plastic emmers met behulp van een zeefdrukmachine (Figuur 3.4). Deze taak vereist zowel technische nauwkeurigheid als aandacht voor veiligheid. In de reële setting is het gebruik van de zeefdrukmachine onderhevig aan slijtage en onderhoudsrisico's, wat



Figuur 3.3: Volledige doosinspectie scene met knop voor het aanvragen van een nieuwe doos (a), knoppen voor aanduiden van “juist” of “fout” (b) en de stapelzone (c).

het bijzonder kostelijk maakt om deze machine enkel voor screeningsdoeleinden in te zetten. Hier biedt VR een duidelijke meerwaarde: de machine hoeft niet te worden aangekocht en gebruikers kunnen de volledige workflow oefenen en doorlopen zonder schade toe te brengen aan machines of materialen.

Voor de virtuele implementatie van deze taak werd een driedimensionaal model van de zeefdrukmachine ontworpen in Blender. Het model werd vervolgens geïmporteerd in de Unity-engine. De machine is voorzien van alle functionele onderdelen die ook in de fysieke wereld terug te vinden zijn, waaronder de invoerzone, drukarm, noodsensoren, start-en noodstop. Het ontwerp werd gebaseerd op observaties van het echte toestel bij Bewel, om een zo realistisch mogelijke ervaring te verkrijgen. Het model is te zien op Figuur 3.5.

De VR-screening start met een korte introductie waarin de gebruiker uitleg krijgt over het gebruik van de machine, de te volgen stappen en de veiligheidsvoorschriften. Vervolgens begint de eigenlijke screening. De eerste stap bestaat erin dat de gebruiker een lege emmer aanvraagt via een knop op het bedieningspaneel (Figuur 3.6a). Wanneer deze knop wordt ingedrukt, verschijnt er een emmer die via een transportband wordt aangeleverd aan de werkpost.

De gebruiker moet de emmer correct oppakken en plaatsen in de drukmodule van de machine (Figuur 3.6b). Dit vergt een correcte positionering, aangezien een foutieve plaatsing kan leiden tot een scheef of onvolledig logo. De machine kan ook worden geactiveerd wanneer er geen emmer in de machine zit, of wanneer deze foutief is aangesloten. Tijdens het drukken dient de gebruiker afstand te houden van de machine: een virtuele nabijheidssensor simuleert een noodstop wanneer de gebruiker te dicht bij de bewegende drukarm komt.

Figuur 3.4: Zeefdruk werkpost bij Bewel.

Na het drukken wordt de gebruiker gevraagd het resultaat te controleren. Het geprojecteerde logo op de emmer kan perfect gecentreerd zijn of kleine afwijkingen vertonen zoals scheefstand, vervaging, of onvolledige bedrukking. De gebruiker moet hierop een beoordeling maken en beslissen of de bedrukking aan de kwaliteitsnorm voldoet.

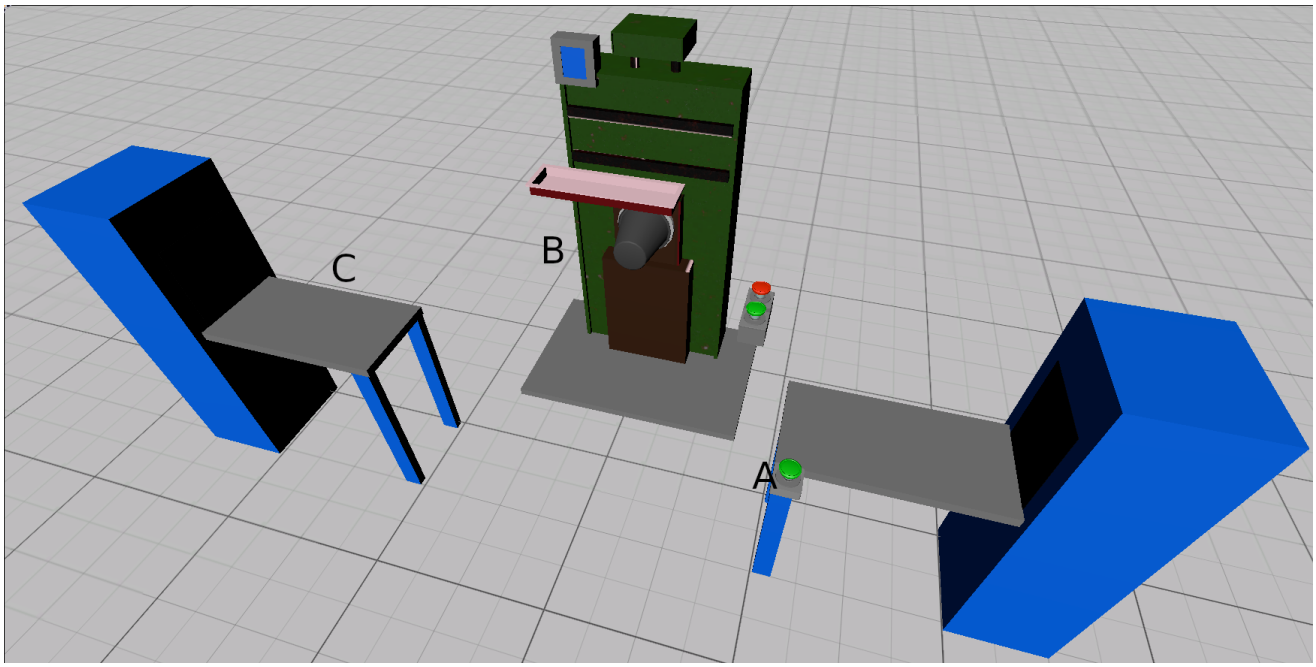
Vervolgens moet de emmer worden verplaatst naar een tweede transportband die leidt naar een virtuele oven (Figuur 3.6c). Deze stap vereist dat de emmer in de correcte oriëntatie wordt geplaatst, namelijk met het logo naar voren gericht.

3.1.5 Emmerstapel-screening

De derde en laatste screening betreft de *emmerstapel-screening*. Deze screening sluit direct aan op het scenario van de zeefdruk screening, en simuleert de handelingen die plaatsvinden nadat een bedrukte emmer uit de oven komt. In de werkvloercontext van Bewel worden bedrukte emmers na het droogproces visueel gecontroleerd op kwaliteit en vervolgens volgens een voorgeschreven patroon gestapeld op een pallet.

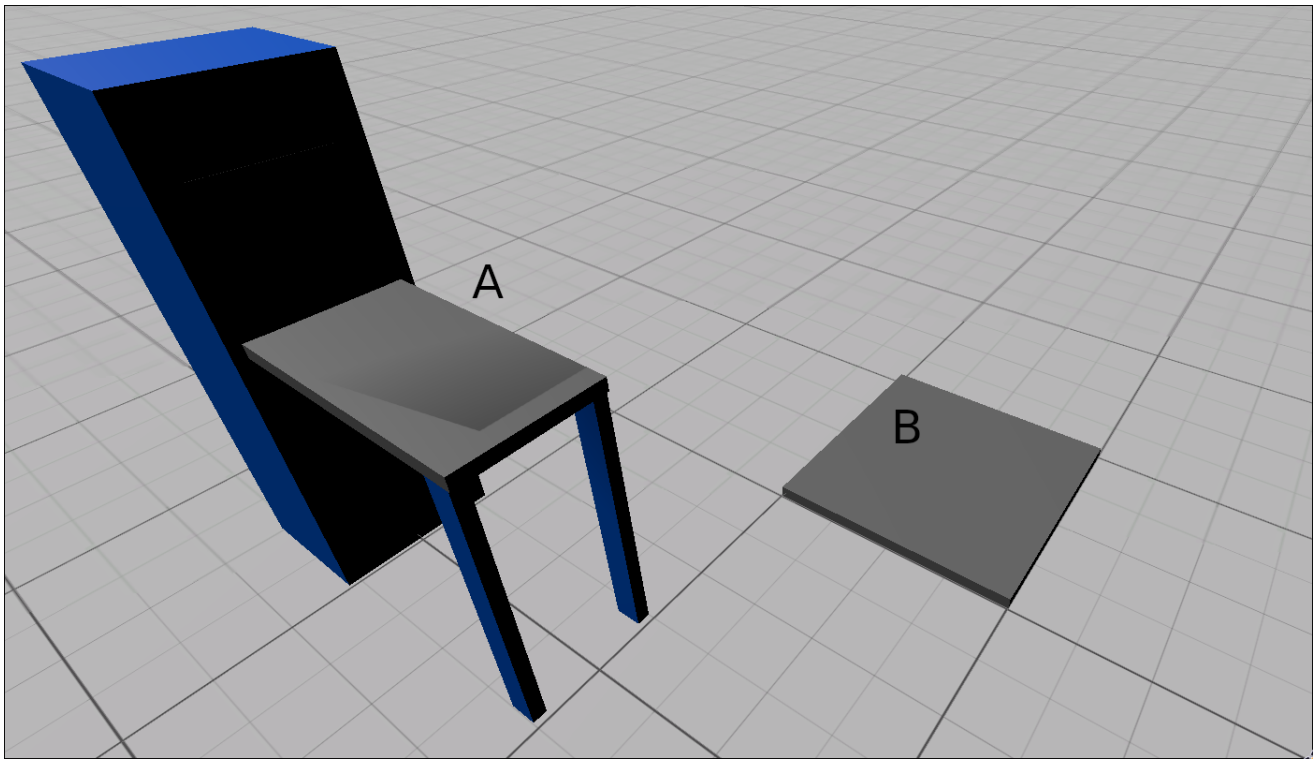
Een eerste cruciale stap is de controle van het logo op de emmer. Aangezien de droogtijd in realiteit varieert, is het mogelijk dat sommige emmers nog nat zijn wanneer ze uit de oven komen. In de virtuele omgeving wordt dit gesimuleerd door de mogelijkheid aan te bieden om met de hand over het logo te vegen. Indien de inkt nog nat is, laat het systeem een visuele verandering van het logo zien (Figuur 3.7), wat aangeeft dat de emmer nog niet klaar is voor stapeling. De gebruiker moet de emmer apart wegleggen.

Figuur 3.5: 3D-model van de zeefdrukmachine.



Figuur 3.6: Volledige zeefdruk scene met knop voor het aanvragen van een nieuwe doos (a), zeefdrukmachine (b) en virtuele oven (c).

Figuur 3.7: Emmer met nat logo voor wrijven (links) en na wrijven (rechts).



Figuur 3.8: Volledige emmerstapel scene met transportband (a) en pallet voor stapelen (b).

Wanneer de emmer als droog en correct geprint wordt beoordeeld, moet deze worden gestapeld op een pallet. Het stapelen moet worden gedaan via een patroon. Het patroon bestaat uit een rasterstructuur, met een vaste volgorde. De gebruiker dient dit patroon correct te interpreteren en de emmers nauwkeurig te plaatsen.

Een bijkomende moeilijkheidsgraad wordt geïntroduceerd via het opgelegde werktempo. Emmers komen met vaste intervallen uit de oven, en de gebruiker krijgt slechts een beperkte tijd om elke emmer te controleren en te stapelen. Indien de gebruiker te traag werkt, hopen emmers zich op en wordt het overzicht moeilijker te behouden. Dit element introduceert tijdsdruk.

3.2 Technologische fundamenten en ontwerpkeuzes

3.2.1 Keuze voor Unity Engine

De applicatie is ontwikkeld met de Unity Engine. Deze keuze is ingegeven door Unity's status als een mature en veelzijdige real-time ontwikkelomgeving, die uitermate geschikt is voor het creëren van complexe en interactieve 3D-toepassingen. Unity biedt een robuuste infrastructuur voor rendering, physics-simulatie en scripting in C#. Bovendien beschikt het over een uitgebreid ecosysteem van tools en een actieve community, wat de ontwikkeling versnelt. De sterke cross-platform ondersteuning van de engine is een bijkomend voordeel, waardoor de applicatie potentieel op diverse hardwareplatformen kan worden ingezet.

3.2.2 Keuze voor OpenXR standaard

Voor de integratie van VR-functionaliteit is gekozen voor de OpenXR-standaard, geïmplementeerd via de officiële Unity-plugin. OpenXR is een open, royalty-vrije standaard van de Khronos Group

die een uniforme API biedt voor VR- en AR-applicaties. Deze keuze is van belang omdat het de applicatie loskoppelt van specifieke hardware. In plaats van te ontwikkelen voor een propriëtaire SDK van één fabrikant, zorgt OpenXR voor compatibiliteit met een breed scala aan huidige en toekomstige VR-headsets. Dit garandeert de duurzaamheid en brede inzetbaarheid van de ontwikkelde screeningtool zonder de noodzaak voor ingrijpende aanpassingen aan de codebasis.

3.2.3 Hardware- en inputmethoden

Als primaire invoermethode is gekozen voor hand-tracking, ondersteund door Unity's XRHands. Deze technologie maakt gebruik van de camera's op de VR-headset om de positie en articulatie van de handen van de gebruiker te detecteren en te vertalen naar een virtuele representatie. Deze keuze is direct gerelateerd aan het doel van het project: het creëren van een intuïtieve en laagdrempelige ervaring, in het bijzonder voor een doelgroep met mogelijk beperkte technologische ervaring. Het gebruik van natuurlijke handbewegingen voor het grijpen en manipuleren van objecten verlaagt de cognitieve drempel aanzienlijk in vergelijking met het aanleren van de knoppenlay-out van traditionele controllers.

Zoals besproken in sectie 2.3, is de ecologische validiteit van een VR-assessment afhankelijk van de mate waarin de simulatie authentiek gedrag uitlokt. Door de gebruiker in staat te stellen met de virtuele ruimte met eigen handen te manipuleren, wordt de virtuele handeling een getrouwere representatie van de overeenkomstige fysieke taak dan bij het gebruik van abstracte controllers. Deze aanpak is tevens een directe toepassing van de inclusieve ontwerpprincipes uit sectie 2.6.4, die essentieel zijn voor de beoogde doelgroep. Door de technologische drempel te verlagen die gepaard gaat met het aanleren van een controller te elimineren, wordt de meting gezuiverd. Dit versterkt de constructvaliditeit van het instrument (sectie 2.8), aangezien de kans wordt geminimaliseerd dat de testresultaten beïnvloed worden door een verstorende variabele.

De XR Interaction Toolkit van Unity dient als de fundamentele laag voor het afhandelen van basisinteracties zoals grijpen en aanraken, waarop de meer gespecialiseerde, op maat gemaakte systemen van dit project zijn gebouwd.

3.3 Architectuur van de VR-scenario's

Een effectieve en betrouwbare virtuele-realiteitssimulatie vereist een architectuur die zowel controleerbaarheid als dynamiek waarborgt. Om reproduceerbare screeningsscenario's te creëren die tegelijkertijd flexibel en uitbreidbaar zijn, werd een spawn-systeem ontwikkeld. Dit systeem vormt de technische ruggengraat voor het gecontroleerd en dynamisch genereren van interactieve objecten.

3.3.1 Dynamisch genereren van objecten

De `SpawnManager`-klasse is de centrale component die verantwoordelijk is voor het dynamisch instantiëren van objecten (zoals dozen of emmers) tijdens een scenario. In plaats van objecten manueel in de Unity-scène te plaatsen, definieert de `SpawnManager` de regels voor hun verschijning. De werking wordt gestuurd door een configureerbare lijst van objecten. Elk `SpawnObjectData`-item specificeert welk object moet worden gecreëerd en, cruciaal, of dit object een fout moet bevatten. Dit mechanisme is de technische basis voor het gecontroleerd introduceren van afwij-

kingen in de screeningtaken, zoals een doos met ontbrekende onderdelen of een verkeerd gekleurd item. De `SpawnManager` beheert verder de locatie, de timing (bv. getimed of op aanvraag) en het maximale aantal te spawnen objecten.

Modulaire configuratie

Om het spawn-proces verder te verfijnen, maakt de `SpawnManager` gebruik van de `ISpawnConfigurator`-interface. Deze interface definieert een contract voor klassen die na-configuratie kunnen toepassen op een zojuist gecreëerd object. Dit volgt het strategy pattern en zorgt voor een zeer flexibele en uitbreidbare architectuur. Enkele concrete implementaties zijn:

- `SpawnConfiguratorRecordable`: Voegt een `TransformRecordableComponent` toe aan het object, waardoor zijn positie en rotatie kunnen worden opgenomen door het data-opnamesysteem.
- `SpawnConfiguratorRandomId`: Wijst een willekeurige identifier toe aan het object, wat essentieel is voor het uniek kunnen volgen van elk item tijdens de analyse.

Door een reeks van deze configuratoren aan de `SpawnManager` te koppelen, kan het gedrag van gespawnde objecten op een modulaire manier worden samengesteld.

Toepassing in de screeningsscenario's

In de praktijk wordt dit systeem ingezet om de screeningscenario's te orkestreren. Voor de doosinspectie-taak wordt de SpawnManager bijvoorbeeld geconfigureerd om een reeks dozen te genereren. Via de `SpawnObjectData`-lijst wordt bepaald dat bijvoorbeeld 70% van de dozen correct is en 30% een specifieke, vooraf gedefinieerde fout bevat. Wanneer de gebruiker op een knop drukt, vraagt deze de `SpawnManager` om het volgende object in de reeks te instantiëren en te configureren. Deze data-gedreven aanpak betekent dat het aanpassen van de moeilijkheidsgraad of het introduceren van nieuwe fouttypes kan gebeuren door enkel de configuratiedata van de SpawnManager aan te passen, zonder de code of de scène zelf te wijzigen. Dit maakt het systeem robuust en eenvoudig te onderhouden en uit te breiden.

3.4 Interactiesysteem voor nauwkeurige manipulatie

Een van de voornaamste technische uitdagingen bij de ontwikkeling van een precieze en betrouwbare screeningtool is de inherente onvoorspelbaarheid van standaard physics-simulaties in combinatie met de beperkte nauwkeurigheid van hand-tracking. Om dit probleem op te lossen, werd een volledig op maat gemaakt interactiesysteem ontwikkeld.

3.4.1 De noodzaak voor een aangepaste interactielaag

In de vroege prototypefase werd voor het stapelen van objecten gebruikgemaakt van de standaard `Rigidbody`-componenten van Unity. Deze aanpak bleek al snel ongeschikt. Kleine, onvermijdelijke trillingen of onnauwkeurigheden in de gedetecteerde handpositie resulteerden frequent in het onbedoeld omstoten, verschuiven of incorrect plaatsen van objecten. Dit leidt niet alleen tot frustratie bij de gebruiker, maar ondermijnt ook de validiteit van de meting zoals besproken in sectie 2.8. Een screening moet de cognitieve en motorische vaardigheid van de gebruiker meten, niet diens vermogen om de beperkingen van een physics engine te compenseren. De standaardaanpak

kon de vereiste gestandaardiseerde en reproduceerbare testomgeving omschreven in sectie 2.3 niet garanderen. Daarom was een meer deterministische methode voor objectplaatsing noodzakelijk.

3.4.2 Bewegingsrestrictie en snapping voor precieze interactie

De kern van de oplossing is de `SlidingSocketInteractor`-klasse, een specialisatie van Unity's `XRSocketInteractor`. In tegenstelling tot een standaard socket die een object onmiddellijk naar een eindpositie teleporteert, functioneert de `SlidingSocketInteractor` als een virtuele glijrail. Het beperkt de beweging van een vastgegrepen object tot een enkele, vooraf gedefinieerde as (X, Y of Z). Wanneer de gebruiker een object in de buurt van de socket brengt, wordt het object aan deze rail 'gekoppeld' en kan het enkel nog lineair langs deze as bewegen.

Daarnaast beschikt de interactor over een configureerbaar snap-gedrag. Zodra het object binnen een bepaalde tolerantie van de doelpositie komt, wordt het automatisch en precies op de correcte positie en rotatie vergrendeld. Dit zorgt voor een interactie die intuïtief aanvoelt voor de gebruiker. De handeling van het plaatsen wordt gerespecteerd terwijl het eindresultaat volledig deterministisch en nauwkeurig is. Het systeem biedt net genoeg hulp om de onnauwkeurigheden in de gedetecteerde handpositie weg te werken, maar niet zo veel dat de gebruiker geen fouten kan maken.

3.4.3 Dynamische generatie van stabiele stapelstructuren

Voortbouwend op de individuele socket, werd de `SlidingSocketStack`-component ontwikkeld om het stabiel stapelen van meerdere objecten te beheren. Wanneer een gebruiker een object succesvol in de onderste socket van een stapel plaatst, gebeuren er twee dingen: het object wordt vergrendeld (zijn `Rigidbody` wordt kinematisch gemaakt om verdere physics-interacties te voorkomen) en de `SlidingSocketStack` instantiëert automatisch een nieuwe `SlidingSocketInteractor` op de correcte positie erboven. Dit proces creëert dynamisch een perfect verticale en stabiele stapel, zonder risico op omvallen.

De `SlidingSocketStackArea` verbreedt dit concept verder naar een groter gebied, zoals een pallet of tafel. Deze component detecteert wanneer een relevant object in zijn zone wordt geplaatst en creëert autonoom een nieuwe `SlidingSocketStack` op die locatie. Dit stelt gebruikers in staat om op meerdere, niet vooraf bepaalde locaties binnen een werkzone te stapelen, wat de flexibiliteit van de scenario's vergroot.

3.4.4 Precisie door interactie filters

Om de robuustheid van de interacties verder te verhogen, werd het systeem uitgebreid met een reeks aangepaste interactiefilters. Deze klassen implementeren Unity's `IXRHoverFilter`- en `IXRSelectFilter`-interfaces en stellen de ontwikkelaar in staat om zeer specifieke voorwaarden te stellen aan interacties.

- `IXROrientationFilter`: Deze filter staat een interactie (zoals het plaatsen in een socket) alleen toe als het object binnen een bepaalde hoektolerantie van een doeloriëntatie is. Dit wordt bijvoorbeeld gebruikt om te verzekeren dat een emmer met de opening in de zeefdrukmachine wordt geplaatst.

- `IXRWhiteListFilter`: Deze filter staat een interactie alleen toe als het object op een expliciete 'witte lijst' staat. Dit wordt bijvoorbeeld gebruikt in de sorteertaak om te garanderen dat enkel dozen op de tafel worden gestapelt, en geen andere objecten die verplaatsbaar zijn in de ruimte.
- `IXRBlackListFilter`: Deze filter blokkeert interactie met objecten op een 'zwarte lijst'. Het wordt onder andere door de `SlidingSocketStack` gebruikt om te voorkomen dat objecten die al deel uitmaken van de stapel per ongeluk opnieuw worden geselecteerd.

Samen zorgt dit systeem ervoor dat de screening een hoge constructvaliditeit heeft. Door de onvoorspelbaarheid van de physics engine te elimineren, wordt de meting gezuiverd. De test evalueert de beslissingen van de gebruiker (bv. welke doos op welke stapel) in plaats van diens behendigheid met een onvoorspelbaar systeem.

3.5 Data-acquisitie, opslag en replay-architectuur

Een fundamentele doelstelling van het project is het vervangen van subjectieve, menselijke observatie door objectieve, data-gedreven analyse. Om dit te realiseren, werd een geavanceerde architectuur voor data-acquisitie en opslag geïmplementeerd. Deze architectuur is specifiek ontworpen voor efficiëntie op standalone VR-headsets, waar rekenkracht en opslagruimte beperkt zijn. Het resultaat is een compleet, objectief en onweerlegbaar verslag van elke gebruikerssessie.

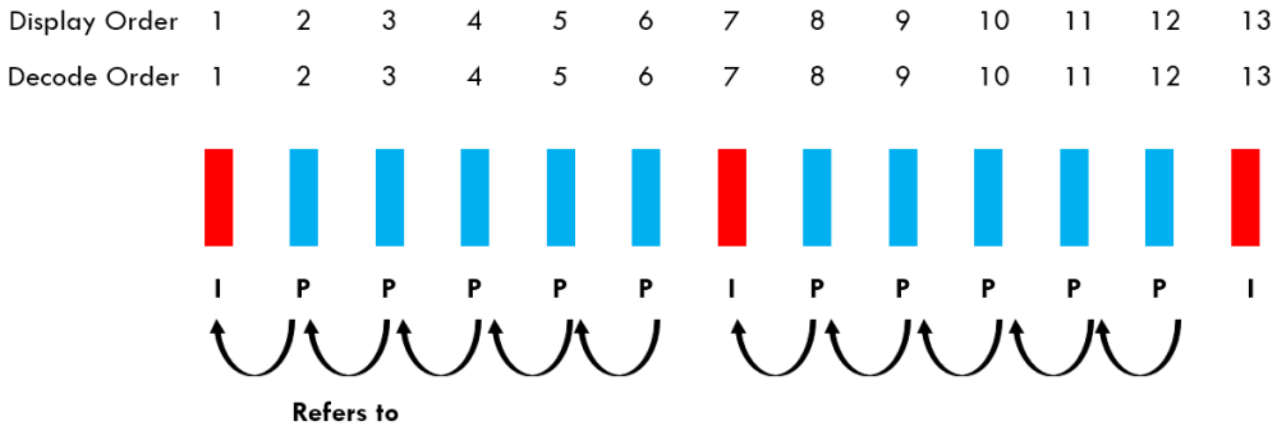
Waar de conventionele screening (sectie 2.1) lijdt onder de beperkingen van een feilbare menselijke observator, vervangt dit systeem die observator door een volledig, onbevooroordeeld en perfect reproduceerbaar digitaal verslag van de prestatie. Deze vastgelegde datastroom bestaande uit posities, rotaties en events met nauwkeurige timestamps, levert de ruwe input voor de objectieve, kwantitatieve prestatiemetingen die in sectie 2.4 worden beschreven, zoals kinematische analyses, tijdmetingen en taakspecifieke foutenanalyses.

3.5.1 Opnemen van de gebruiker

De `RecordingManager`-klasse is verantwoordelijk voor het vastleggen van de toestand van de simulatie tijdens een sessie. In plaats van de volledige toestand van alle objecten bij elke frame op te slaan, een methode die te veel rekenkracht en dataverkeer zou vergen, is een efficiëntere aanpak geïmplementeerd, geïnspireerd op videocompressietechnieken. Het systeem maakt onderscheid tussen twee typen frames (Figuur 3.9) [19]:

- **Intra-coded frames (I-frames):** Een I-frame is een volledige 'snapshot' van de toestand van alle relevante, opneembare objecten in de scène op een specifiek tijdstip. Dit omvat posities, rotaties en andere specifieke statussen.
- **Predictive-coded frames (P-frames):** Een P-frame slaat enkel de toestand van de veranderde object op ten opzichte van het voorgaande frame. Bijvoorbeeld, als een object niet is bewogen, wordt er geen data voor opgeslagen in het P-frame.

Deze hybride aanpak, beheerd door de `RecordingManager`, verlaagt de hoeveelheid opgeslagen data drastisch, terwijl een hoge temporele resolutie behouden blijft. Dit maakt het mogelijk om gedetailleerde sessies op te nemen zonder de prestaties van de VR-applicatie negatief te beïnvloeden, wat cruciaal is op standalone hardware. Objecten die moeten worden opgenomen,



Figuur 3.9: Visualisatie van I- en P-frames.

implementeren de IRecordable-interface, wat het systeem modulair en eenvoudig uitbreidbaar maakt.

3.5.2 Reconstrueren van de opgenomen data

De opgenomen data wordt gebruikt door de `ReplayManager` om een sessie volledig en interactief te reconstrueren. Deze component leest de opgeslagen frame-data en past de toestanden van de objecten in de scène aan om de originele sessie na te bootsen. De `ReplayManager` is de motor achter de simulatieweergave in het resultatendashboard en ondersteunt functionaliteiten zoals afspelen, pauzeren, de afspeelsnelheid aanpassen en direct naar een specifiek tijdstip in de opname springen. Bij het navigeren naar een tijdstip zoekt de manager het dichtstbijzijnde voorgaande I-frame en past vervolgens alle tussenliggende P-frames toe om de scène efficiënt te reconstrueren. Objecten die niet aanwezig zijn in het geselecteerde I-frame worden verwijderd, en nieuwe objecten worden gecreëerd op basis van de gegevens in het frame.

Deze aanpak minimaliseert de hoeveelheid werk die nodig is om de sc‘ene te reconstrueren, terwijl de nauwkeurigheid van de replay behouden blijft. Het gebruik van I-frames als referentiepunten maakt het navigeren efficiënt, zelfs in lange opnames.

3.5.3 Persistente en gecomprimeerde opslag

Aan het einde van elke screeningsessie worden alle verzamelde gegevens persistent opgeslagen op het apparaat. De `SaveDataHandler`-klasse orkestreert dit proces. Het verzamelt data van verschillende bronnen die de `ISaveData`-interface implementeren, waaronder de opnamedata van de RecordingManager, de event-logs met alle gebruikersinteracties, en de resultaten van de validatiemodules. Al deze data wordt geserialiseerd naar een gestructureerd JSON-formaat. Om de opslagruimte op de headset te minimaliseren, wordt dit JSON-bestand vervolgens automatisch gecomprimeerd in een .zip-archief. Deze gecomprimeerde bestanden kunnen later eenvoudig van het apparaat worden gehaald voor externe analyse of back-up.

3.6 Geautomatiseerd beoordelingsmodel

De objectieve data die door het opnamesysteem wordt verzameld, vormt de input voor een geautomatiseerd beoordelingsmodel. Dit model is ontworpen met modulariteit en flexibiliteit als kernprincipes, waardoor het mogelijk is om de beoordelingscriteria nauwkeurig af te stemmen op de specifieke competenties die men wenst te meten.

3.6.1 Modulaire validatiecomponenten

De validatielogica is opgesplitst in gespecialiseerde componenten, conform het Separation of Concerns-principe. Elke validator is verantwoordelijk voor een specifiek aspect van het eindresultaat.

- `SortValidator`: Deze component analyseert de sortering van objecten. Het controleert of objecten correct zijn gegroepeerd op basis van een eigenschap (bv. kleur) en of er geen items van een verkeerd type binnen een groep aanwezig zijn. Een typisch voorbeeld is het detecteren van een blauwe doos in een stapel die voor rode dozen bestemd is.
- `StackValidator`: Deze component focust op de fysieke correctheid van een stapel. Het valideert aspecten zoals de onderlinge afstand tussen objecten, de correcte rotatie, en of de maximale stapelhoogte niet is overschreden.

Deze modulaire aanpak maakt het mogelijk om verschillende soorten fouten onafhankelijk van elkaar te detecteren en te kwantificeren.

3.6.2 Van ruwe data naar competentiescore

De `ScoringValidator` is de centrale motor van het beoordelingssysteem. Deze klasse fungeert als een aggregator die data uit meerdere bronnen consumeert:

- De resultaten van de `SortValidator` en `StackValidator`.
- De ruwe event-logs die alle gebruikersinteracties bevatten (bv. grijpen, plaatsen, knopdrukken).
- Specifieke inspectiedata (bv. of een doos correct als 'juist' of 'fout' is gelabeld).

De `ScoringValidator` verwerkt deze input en transformeert deze naar een lijst van kwantitatieve datapunten. Voorbeelden hiervan zijn voltooiingstijd, aantal fouten en het totaal aantal handelingen. Deze datapunten vormen de objectieve, numerieke representatie van de prestatie van de gebruiker.

3.6.3 Een flexibel en regelgebaseerd Systeem

De kracht van het beoordelingsmodel ligt in zijn flexibiliteit, die wordt gerealiseerd door een op regels gebaseerd systeem. De `ScoringValidator` past een configureerbare lijst van `ScoringRuleData`-objecten toe om de uiteindelijke score te berekenen. Elke ScoringRuleData definieert een eenvoudige, maar krachtige regel: het koppelt een datapunt aan een voorwaarde (bv. 'groter dan', 'gelijk aan') en een drempelwaarde. Als aan de voorwaarde is voldaan, wordt een gespecificeerde score toegekend (positief of negatief).

De keuze voor een regelgebaseerd systeem is een bewuste ontwerpafweging die prioriteit geeft aan transparantie en gebruikersvertrouwen. Hoewel de literatuur (sectie 2.4.2) de potentie van complexere machine learning modellen zoals ANN of HMM aantoont, functioneren dergelijke modellen vaak als een 'black box'. Het is voor een begeleider moeilijk of onmogelijk om te doorgronden hoe een dergelijk model tot een specifieke score komt. Een expliciet, regelgebaseerd systeem is daarentegen een 'glass box': elke component van de score kan direct worden herleid tot een observeerbare gebeurtenis en een begrijpelijke regel. Het stelt de begeleider in staat de logica van het systeem te valideren, te controleren en indien nodig aan te passen.

3.7 Resultatendashboard voor begeleiders

Om de objectieve data en de berekende scores om te zetten in bruikbare inzichten voor begeleiders, werd een resultatendashboard ontwikkeld. Dit dashboard fungeert als een geïntegreerde analyseomgeving die verschillende perspectieven op de prestatie van een gebruiker combineert. Het ontwerp is gericht op transparantie en interactiviteit, waardoor het systeem niet als een vervanging dient, maar als een hulpmiddel dat de begeleider in staat stelt een objectievere en consistentere beoordeling te maken.

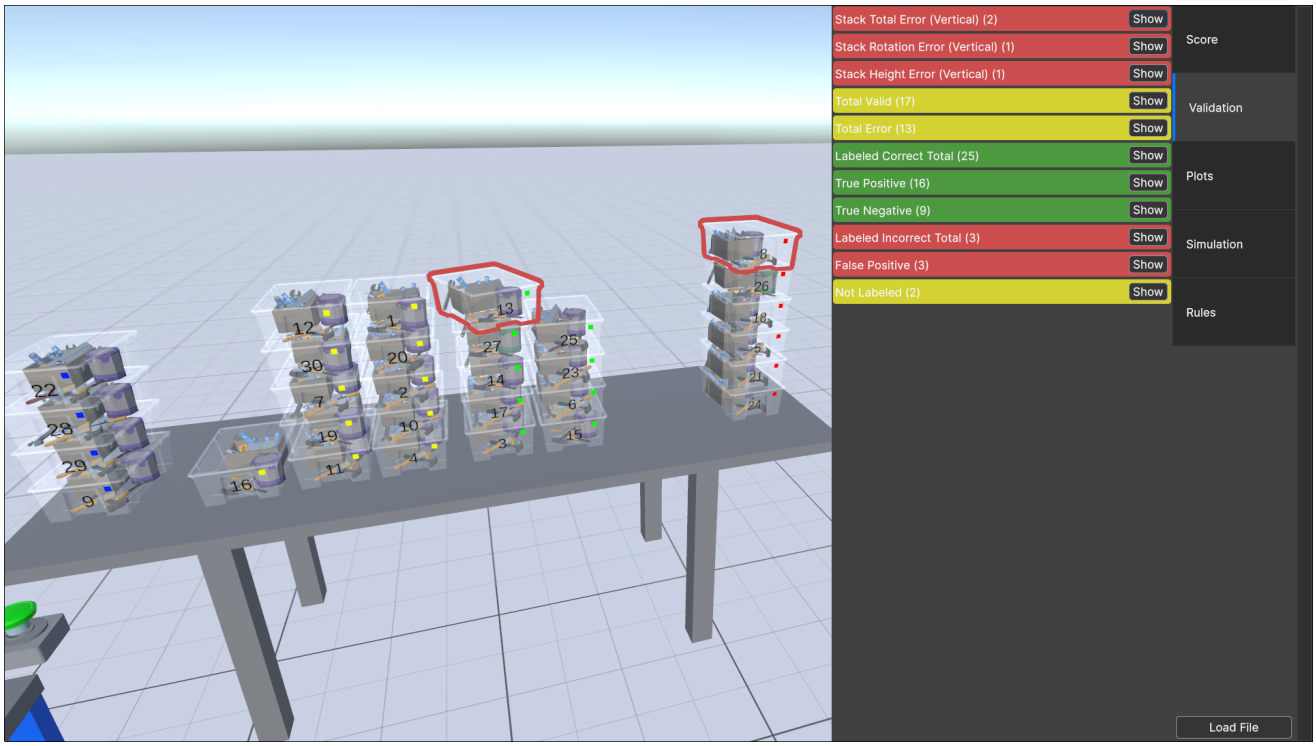
3.7.1 Geïntegreerde analyseomgeving

Het dashboard presenteert de resultaten van een screeningsessie via verschillende, met elkaar verbonden modules. Deze architectuur stelt de begeleider in staat om een prestatie vanuit meerdere invalshoeken te analyseren.

Een eerste analyse start bij de interactieve 3D-resultaatweergave. Dit component toont het eindresultaat van de uitgevoerde taak, zoals de opgestapelde dozen, in een navigeerbare 3D-scène. Fouten die door de validatiemodules zijn gedetecteerd, zoals incorrect gesorteerde of geplaatste objecten, worden hierin direct visueel gemarkeerd, bijvoorbeeld door de betreffende objecten rood te kleuren. Deze weergave, te zien in Figuur 3.10, biedt een onmiddellijk en intuïtief overzicht van het eindproduct. Het dient als een efficiënt startpunt voor de analyse, vergelijkbaar met de productgebaseerde metingen uit conventionele screenings, maar verrijkt met geautomatiseerde en objectieve foutdetectie.

Vervolgens kan de analyse verdiept worden met de 2D-datavisualisaties. Deze module vertaalt de ruwe data van de sessie naar overzichtelijke grafieken en diagrammen, zoals te zien in Figuur 3.11. Kwantitatieve datapunten die door de `ScoringValidator` zijn gegenereerd, zoals de prestatiesnelheid over tijd, de frequentie van specifieke fouttypes of de verdeling van de tijd over verschillende subtaken, worden hier gevisualiseerd. Deze weergave verplaatst de analyse van het ruimtelijke naar tijd domein. Het stelt de begeleider in staat om patronen en trends te identificeren die onzichtbaar zijn in het eindresultaat, zoals een geleidelijke afname in werktempo of een terugkerende aarzeling bij een specifieke handeling.

De meest diepgaande analyse wordt mogelijk gemaakt door de interactieve simulatieweergave. Dit centrale onderdeel van het dashboard biedt een volledige, interactieve 3D-replay van de gebruikerssessie, aangedreven door de `ReplayManager` (zie Figuur 3.12). De begeleider kan de sessie vanuit meerdere camerastandpunten bekijken, het afspelen pauzeren, de snelheid aanpassen en direct navigeren naar specifieke momenten in de tijdlijn. Deze functionaliteit levert de context



Figuur 3.10: Interactieve 3D-weergave van het eindresultaat in het resultatendashboard.

die abstracte data en statische eindresultaten missen. Een begeleider kan bijvoorbeeld vaststellen of een fout het gevolg was van onoplettendheid, een misinterpretatie van de instructies, of een motorische onhandigheid.

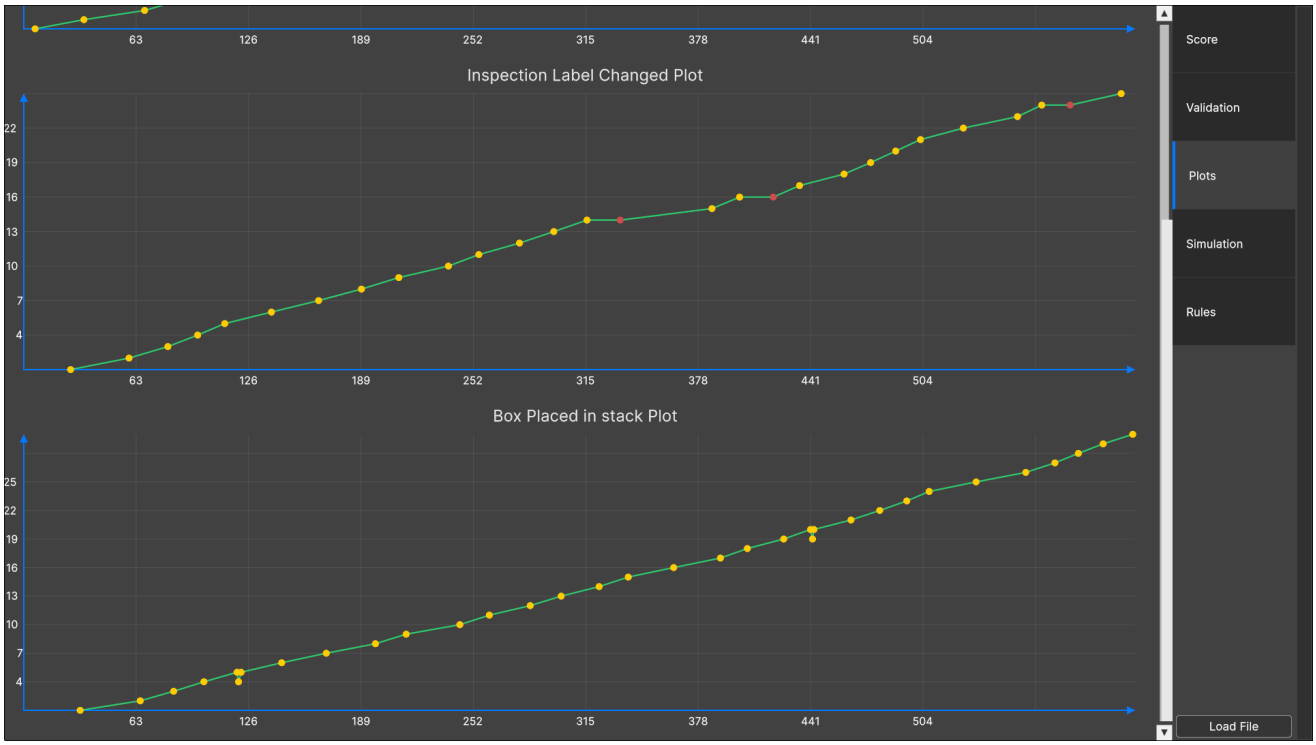
3.7.2 Transparante kwantificering en configuratie

Naast de visuele analyse biedt het dashboard krachtige tools om de prestatie te kwantificeren en de beoordelingslogica zelf te beheren. Deze componenten zijn ontworpen volgens een 'glass box'-filosofie, die prioriteit geeft aan transparantie en controle voor de begeleider.

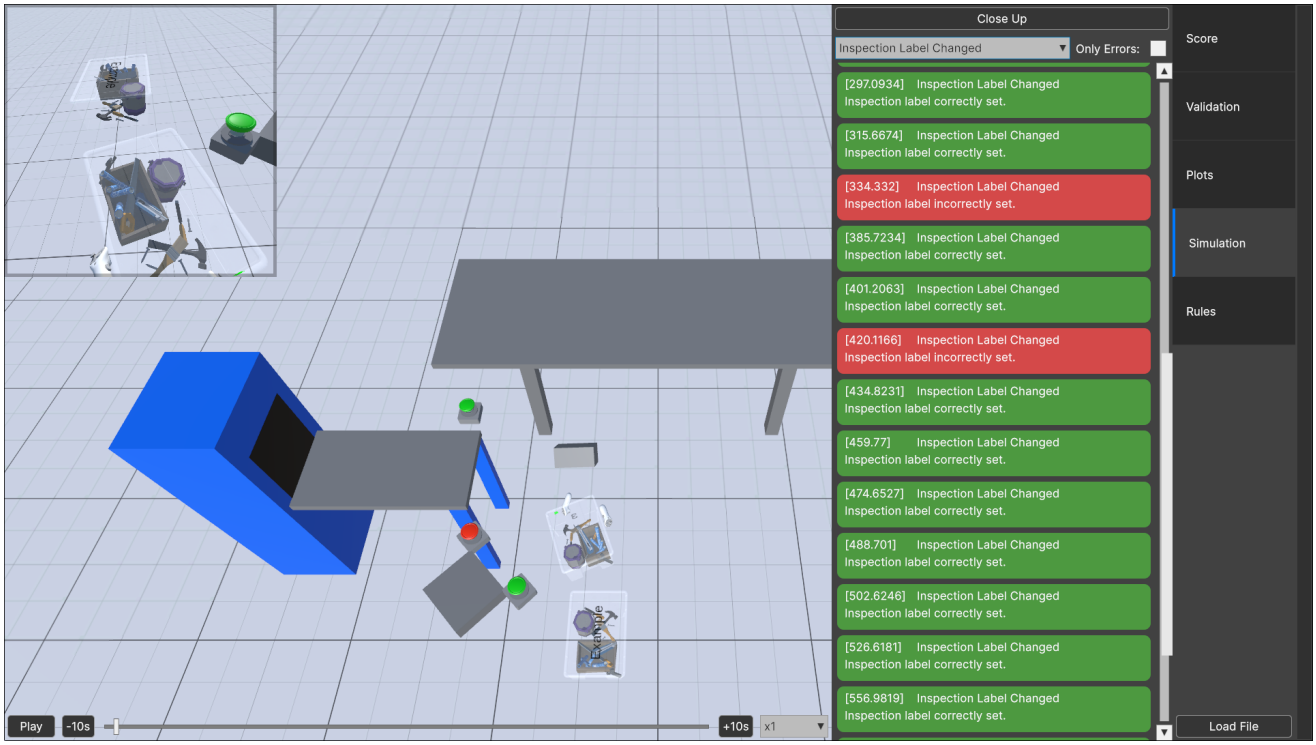
De scorevisualisatie visualiseert de score die door de `ScoringValidator` wordt bekomen. Zoals getoond in Figuur 3.13, presenteert deze module niet simpelweg een eindcijfer, maar vertaalt de abstracte, kwantitatieve datapunten die door de `ScoringValidator` zijn verwerkt naar de kwalitatieve, competentiegerichte taal die binnen de organisatie wordt gebruikt. Deze vertaalslag is belangrijk om betekenisvolle, concrete en actiegerichte feedback te kunnen geven. Het stelt een begeleider in staat om een specifieke competentiescore direct te koppelen aan het objectieve bewijs uit de datavisualisaties en de simulatieweergave, waardoor het feedbackgesprek volledig data-onderbouwd wordt.

De ultieme controle over het beoordelingsproces wordt geboden door de editor voor beoordelingsregels. Dit component, te zien in Figuur 3.14, legt de volledige logica van het ScoringValidator-systeem bloot en maakt deze configureerbaar. Dit systeem geeft de begeleider de mogelijkheid om de beoordelingscriteria volledig te beheren. Een begeleider kan bijvoorbeeld individuele regels activeren of deactiveren, de drempelwaarden van een regel aanpassen of de toegekende score van een regel wijzigen.

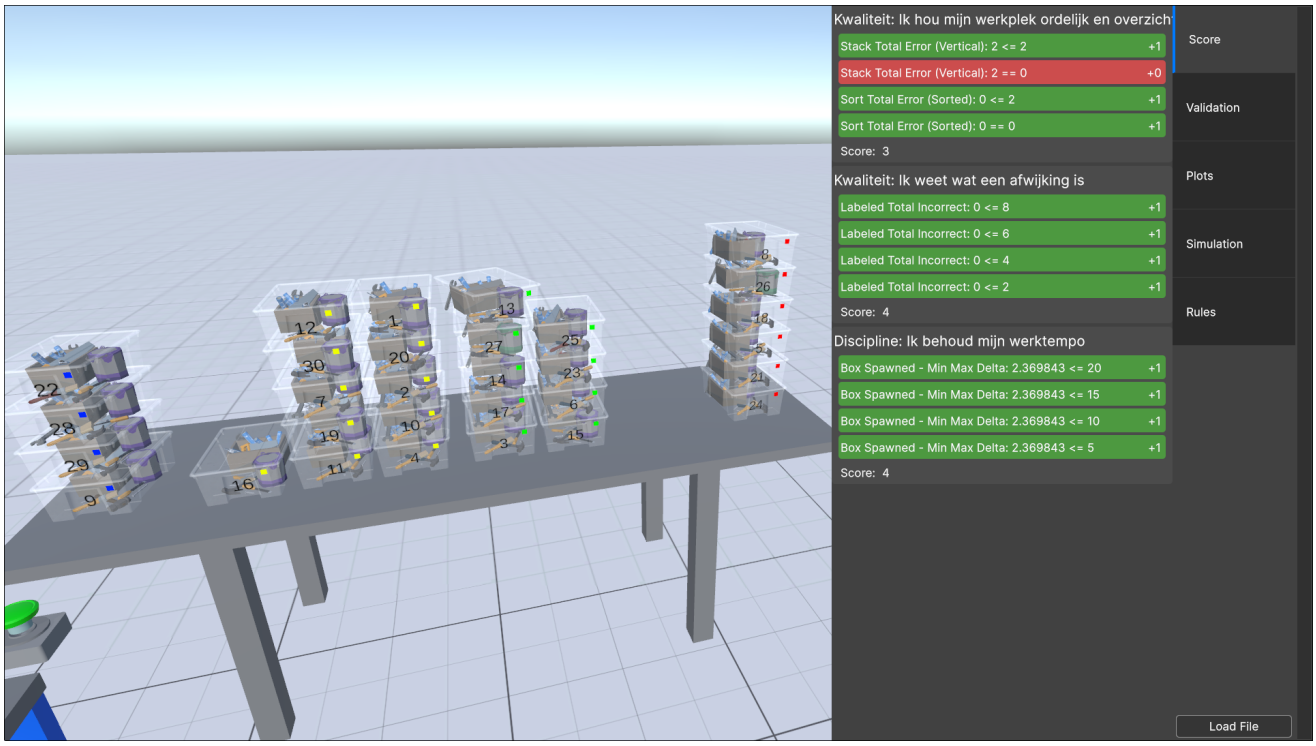
Het ontwerp van het dashboard is hierbij een direct antwoord op de beperkingen van conventi-



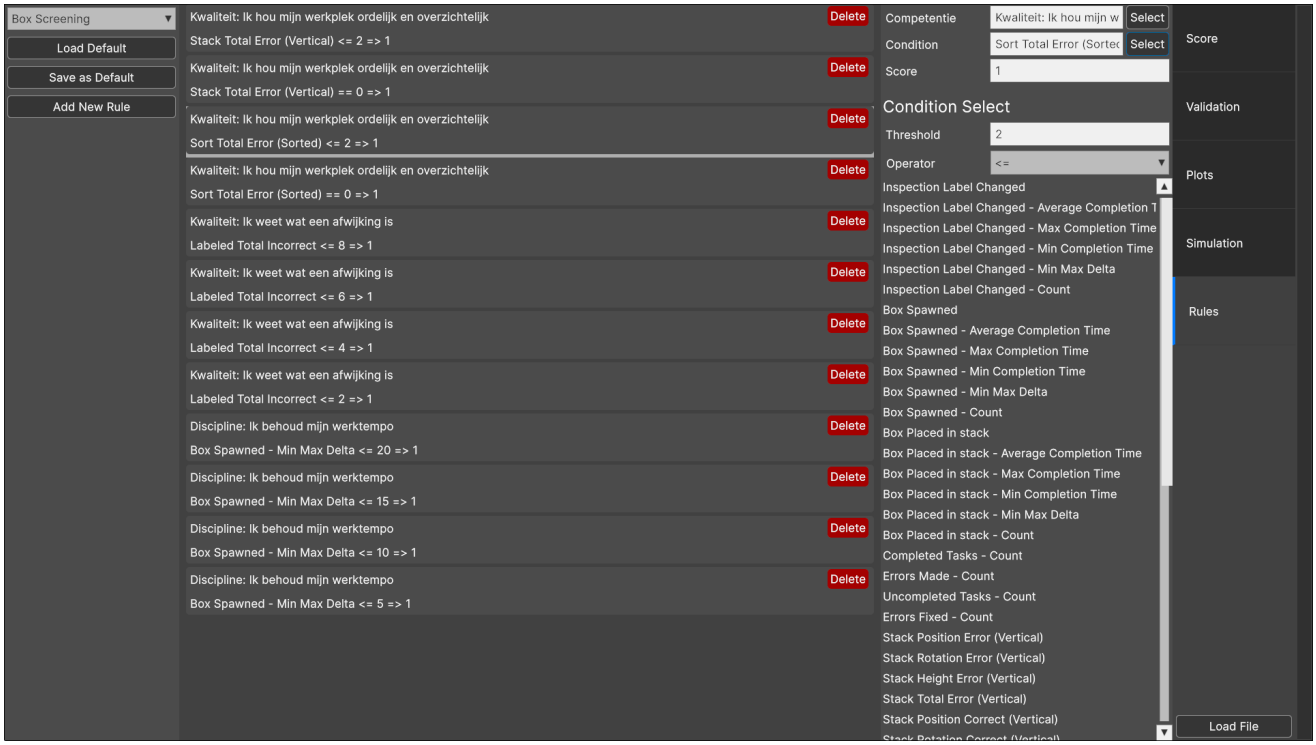
Figuur 3.11: 2D-Datavisualisaties in het resultatendashboard.



Figuur 3.12: Interactieve 3D-replay in het resultatendashboard.



Figuur 3.13: Score weergave in het resultatendashboard.



Figuur 3.14: beoordelingsregels editor in het resultatendashboard.

onele screening die in sectie 2.1 werden geïdentificeerd. De verschillende componenten van het dashboard zijn specifiek ontworpen om de fundamentele zwaktes van menselijke observatie te compenseren. De 2D-datavisualisaties (Figuur 3.11) bestrijden de inherente subjectiviteit door objectieve, kwantitatieve data te presenteren. De simulatieweergave (Figuur 3.12) lost het gebrek aan procesinzicht op door de begeleider in staat te stellen niet alleen het eindresultaat te beoordelen, maar het volledige proces te observeren en te analyseren hoe en waarom bepaalde handelingen werden uitgevoerd. Het dashboard functioneert zodoende als een cognitief hulpmiddel dat, in lijn met de literatuur (sectie 2.2), de menselijke expertise van de begeleider niet vervangt, maar versterkt met objectieve en gedetailleerde data, waardoor deze zich kan richten op interpretatie, contextualisering en het geven van gerichte feedback.

Hoofdstuk 4

Evaluatie

4.1 VR-screening

Deze initiële studie richtte zich op een kleine, maar representatieve groep van vier maatwerkers. Het hoofddoel van deze studie was niet het meten van prestaties, maar het evalueren van de ontwikkelde VR screeningstoepassing op drie vlakken: de bruikbaarheid van de software, de subjectieve beleving door de eindgebruiker en de eventuele aanwezigheid van ongewenste fysieke effecten, zoals duizeligheid of misselijkheid.

De studie maakte gebruik van een Meta Quest 3 headset, een moderne, standalone VR-bril die geen externe computer of kabels vereist. Deze keuze draagt bij aan een laagdrempelige en veilige ervaring. Als primaire invoermethode werd gekozen voorhand-tracking.

Elke testsessie met een deelnemer volgde een vierstaps protocol:

- **Introductie**: De begeleider legde het doel van de test uit. Hierbij werd expliciet benadrukt dat de software en de VR-toepassing werden getest, en niet de vaardigheden van de deelnemer. Deze framing is een beproefde techniek uit gebruikersonderzoek om prestatieangst te verminderen en natuurlijk gedrag te stimuleren.

- **Instructies**: De deelnemer ontving een korte, duidelijke uitleg over het gebruik van de headset en de specifieke regels van de uit te voeren VR-taak.

- **Screening in VR**: De deelnemer voerde de taak zelfstandig uit in de virtuele omgeving. Een begeleider bleef gedurende de hele sessie beschikbaar voor ondersteuning en om in te grijpen indien nodig. Deelnemers werden ook geïnformeerd dat zij op elk moment konden pauzeren of stoppen als zij zich onwel voelden.

- **Post-test vragenlijst**: Onmiddellijk na de VR-ervaring werd een gestructureerde vragenlijst afgenomen.

De virtuele taak die de deelnemers moesten uitvoeren, was de doosinspectie-screening. Het betrof een gesimuleerde werkpost voor kwaliteitscontrole en sortering die een combinatie van verschillende competenties vereist. Deelnemers moesten in totaal 10 dozen verwerken.

4.2 Resultatendashboard voor begeleiders

4.2.1 Methodologie

De evaluatie van de VR-gebaseerde screeningstool werd uitgevoerd aan de hand van een pilootstudie, ontworpen om de effectiviteit en bruikbaarheid van het systeem te meten. De aanpak was gericht op het verkrijgen van inzicht in drie centrale onderzoeksvragen:

- RQ1: De bruikbaarheid en begrijpelijkheid van het dashboard.
- RQ2: In hoeverre de VR-tool bijdraagt aan een objectievere evaluatie.
- RQ3: De invloed van de toegevoegde digitale informatie, zoals datavisualisaties (plots) en een 3D-simulatie, op de uiteindelijke beoordeling.

De studie werd uitgevoerd in twee fasen. De eerste fase omvatte de bruikbaarheid en validatie van het testsysteem. Dit gebeurde in samenwerking met een expert van Bewel.

De tweede fase bestond uit een gebruikersonderzoek met 12 deelnemers. Tijdens dit onderzoek kregen de deelnemers via een dashboard zes verschillende, vooraf opgenomen Virtual Reality (VR) sessies voorgelegd. Voorafgaand aan de test werd een zevende introductiesessie aangeboden. Deze oefensessie had als doel de deelnemers vertrouwd te maken met de VR-omgeving en de besturing. De gegevens uit deze introductiesessie werden niet meegenomen in de uiteindelijke data-analyse.

De zes VR-sessies (A-F) zijn ontworpen om specifieke, herkenbare gedragsprofielen te simuleren die relevant zijn voor de sorteertaak. Een overzicht van deze profielen wordt gegeven in tabel 4.1.

Tabel 4.1: Omschrijving van de gesimuleerde gedragsprofielen

A	Alles juist: correct sorteren en controleren van alle potjes gedurende de volledige simulatie.
B	Niet opmerkzaam: fouten ontstaan doordat foutieve potjes toch als correct worden gelabeld.
C	Te kritisch: correcte potjes worden onterecht als fout gemarkeerd.
D	Afgeleid: start met een goed tempo, maar steeds trager naarmate de tijd vordert. De persoon is vaak afgeleid.
E	Afglijdend gedrag: de persoon begint sterk, maar maakt geleidelijk steeds meer fouten zonder herstel.
F	Foute kleurstapeling: bij het stapelen worden kleuren niet gegroepeerd, maar bijvoorbeeld afwisselend gestapeld.
Voorbeeld	Het scenario start met veel fouten, maar gaandeweg herpakt de persoon zich en worden er minder fouten gemaakt.

Om de invloed van volgorde-effecten te minimaliseren, waarbij de volgorde waarin taken worden aangeboden het gedrag of de prestatie van de deelnemer kan beïnvloeden, werd er gebruik gemaakt van een *balanced Latin square*-ontwerp. Dit statistische ontwerp garandeert dat elke VR-sessie (A–F) even vaak op elke positie in de testvolgorde (1 t/m 6) verschijnt. Bovendien zorgt het ervoor dat elke sessie even vaak voorafgegaan wordt door elke andere sessie. Op deze wijze worden zowel directe volgorde-effecten als overgangseffecten systematisch over de deelnemers verdeeld.

De toegepaste balanced Latin square voor de zes gebruikte sessies is weergegeven in tabel 4.2. Elke rij vertegenwoordigt de volgorde van sessies die een deelnemer te zien kreeg. Aangezien er 12 deelnemers waren, werd deze matrix tweemaal doorlopen.

Tabel 4.2: Balanced Latin Square voor zes VR-sessies (A–F)

	1	2	3	4	5	6
1	A	B	F	C	E	D
2	B	C	A	D	F	E
3	C	D	B	E	A	F
4	D	E	C	F	B	A
5	E	F	D	A	C	B
6	F	A	E	B	D	C

De evaluatie was gestructureerd als een incrementele test, waarbij informatie stapsgewijs werd onthuld om de invloed van elke component te kunnen isoleren. Na afloop van alle sessies vulden de deelnemers een vragenlijst in. Deze dataverzameling had het karakter van een semi-gestructureerd interview, waarin een vaste set vragen op een Likertschaal werd gecombineerd met een reeks mondelinge, open vragen voor kwalitatieve feedback.

4.2.2 Deelnemers

De evaluatie werd in twee fasen uitgevoerd. De eerste test werd afgenomen bij één expert van Bewel, een persoon met ervaring in het screenen van de doelgroep.

Een daaropvolgende, uitgebreidere studie werd uitgevoerd met een groep van 12 deelnemers. Deze groep bestond uit 11 mannen en 1 vrouw, met een leeftijd tussen 21 en 30 jaar.

4.2.3 Procedure

Deelnemers doorliepen een gestructureerd testprotocol met behulp van een dashboard waarin vooraf opgenomen VR-sessies werden geanalyseerd. De test omvatte de evaluatie van zes verschillende scenario's, die elk een specifiek gedragspatroon representeerden, zoals foutloos werken, afnemende prestaties of herstelgedrag 4.1. Een zevende scenario werd als initiële casus gebruikt om de deelnemers met het systeem vertrouwd te maken.

Voor elke sessie werd de evaluatie opgedeeld in vier fasen, waarbij stelselmatig meer informatie werd vrijgegeven:

- Enkel eindresultaat: De deelnemer beoordeelde de prestatie uitsluitend op basis van het visuele eindresultaat van de gesorteerde en gestapelde items (Figuur 3.10).
- Toevoegen van data plots: Grafieken van onder meer gemaakte fouten, uitvoeringstijd en interacties werden aan het dashboard toegevoegd (Figuur 3.11).
- Toevoegen van simulatie: Een volledige 3D-herhaling van de VR-sessie werd beschikbaar gesteld, waardoor de deelnemer het gedrag van de gescreende persoon kon observeren (Figuur 3.12).
- Tonen van systeemscore: Tot slot werd de door het algoritme berekende score getoond (Figuur 3.13).

Na elke fase werd aan de deelnemers gevraagd om een score tussen de 1 en 4 te geven op 3 verschillende criteria, samen met een zekerheidsscore op een schaal van 1 tot 5. De 3 criteria zijn:

- Orde en overzicht wordt bewaard tijdens het uitvoeren van de taak. Het afgeleverd resultaat is overzichtelijk voor de monitor.
- Ik zie het verschil tussen het voorbeeld en de inhoud van de dozen. Ik duid dit correct aan.
- Werk wordt binnen de vooropgestelde tijd afgewerkt. Er is een regelmatig werkritme zonder lange pauzes of vertraging.

De schaal van 1 tot 4 voor deze beoordelingscriteria is gekozen omdat dit aansluit bij de evaluatiemethodiek die Bewel hanteert. De huidige werkwijze bij Bewel komt overeen met enkel de eerste fase van deze procedure, waarbij uitsluitend het eindresultaat van het werk wordt gecontroleerd.

Na afloop van alle sessies vulden de deelnemers een post-test vragenlijst in met zowel gesloten Likert-schaalvragen als open vragen om diepgaandere feedback te verzamelen.

4.2.4 Gegevensanalyse

De verzamelde data werden geanalyseerd met een combinatie van kwantitatieve en kwalitatieve methoden. De kwantitatieve analyse richtte zich op de scores en zekerheidsniveaus die in de verschillende fasen van de procedure werden genoteerd. Hierbij werd onderzocht of en hoe de beoordeling van de deelnemers veranderde naarmate meer informatie (plots, simulatie, systeemscore) beschikbaar kwam.

De kwalitatieve analyse bestond uit het thematisch coderen van de antwoorden op de open vragen uit de post-test vragenlijst. Deze analyse was gericht op het identificeren van de meest en minst nuttige onderdelen van het dashboard, het begrijpen van de redenering achter de beoordelingen en het verzamelen van suggesties voor verdere verbetering van het systeem. De feedback op de Likert-schaalvragen werd eveneens geanalyseerd om de algemene perceptie van bruikbaarheid, vertrouwen en objectiviteit te meten.

Hoofdstuk 5

Resultaten

5.1 VR-screening

Dit hoofdstuk presenteert de resultaten van de gebruikersevaluatie van de ontwikkelde VR-screeningapplicatie. De evaluatie had tot doel de bruikbaarheid, gebruikerservaring en algehele aanvaarding van de VR-tool te beoordelen door de beoogde eindgebruikers. Hiertoe werd de applicatie getest door een representatieve groep van vier maatwerkers van de organisatie Bewel.

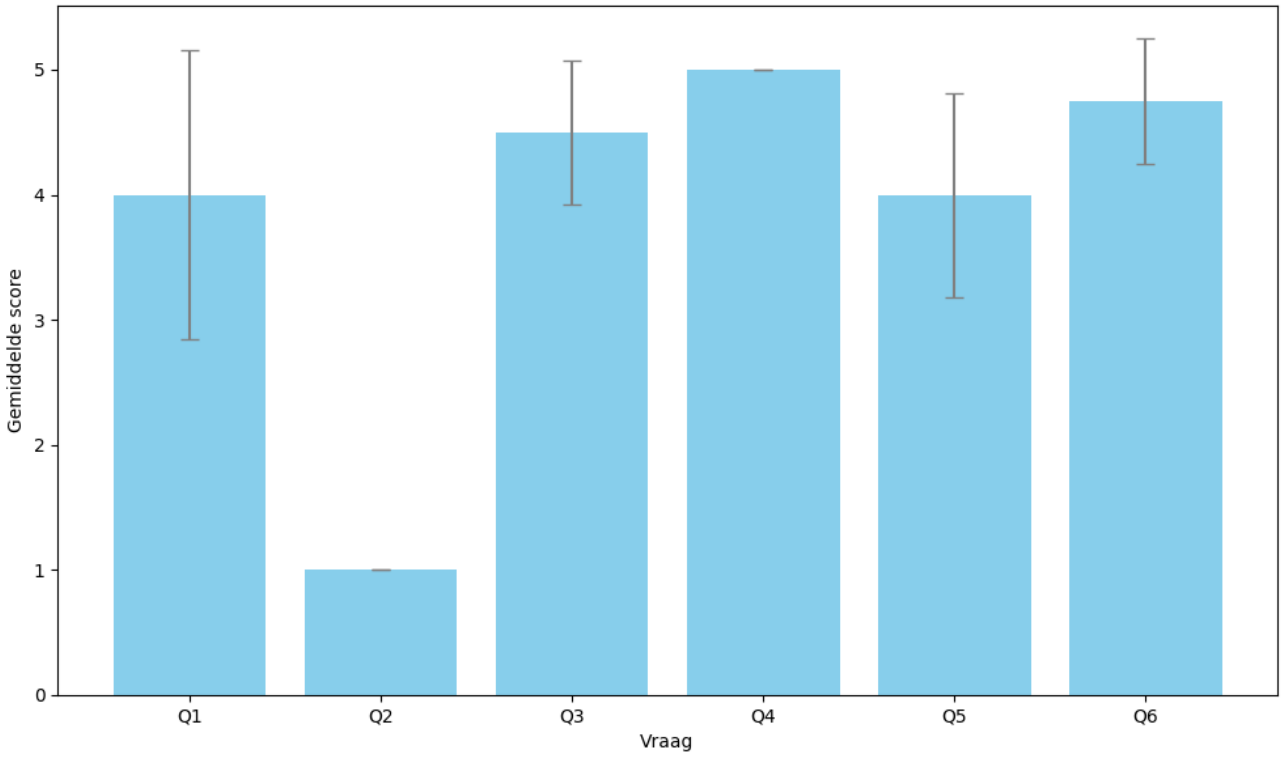
5.1.1 Analyse van de gebruikerservaring

De kwantitatieve data, verzameld via de Likert-schaalvragen, bieden een numeriek onderbouwd inzicht in de onmiddellijke reacties van de deelnemers op de VR-screening. De resultaten, samengevat in Figuur 5.1, wijzen op een overwegend tot zeer positieve gebruikerservaring over de gehele lijn. Scores werden gegeven op een 5-puntsschaal, waarbij 1 staat voor 'helemaal niet akkoord' / 'geen last' en 5 voor 'helemaal akkoord' / 'zeer veel last'.

De analyse van deze scores onthult verschillende bevindingen. Ten eerste werd de helderheid van de instructies (Figuur 5.1 Q4) unaniem als perfect beoordeeld, met een gemiddelde score van M=5.0 en een standaardafwijking van SD=0.0. Dit resultaat is van belang, omdat het aangeeft dat de deelnemers de taak en de bediening van de applicatie volledig begrepen alvorens te starten, wat een voorwaarde is voor een valide competentiemeting. Het gebruiksgemak (Figuur 5.1 Q5) van de software zelf werd eveneens als hoog beoordeeld (M=4.0, SD=0.82), wat duidt op een intuïtief ontwerp en een lage drempel voor de gebruikers.

Een bijzonder opmerkelijk resultaat betreft het fysieke comfort van de deelnemers. De vraag naar de aanwezigheid van duizeligheid of misselijkheid (Figuur 5.1 Q2), symptomen die gezamenlijk bekendstaan als cybersickness, leverde een gemiddelde score van M=1.0 op, met een standaardafwijking van SD=0.0. Deze score geeft aan dat elke deelnemer aangaf geen enkele last te hebben ondervonden.

Voortbouwend op dit fundament van comfort en duidelijkheid, blijkt de affectieve respons van de deelnemers uitzonderlijk positief. De ervaring werd als zeer leuk beoordeeld (Figuur 5.1 Q3) (M=4.5, SD=0.58), wat wijst op een hoge mate van gebruikerstevredenheid en engagement. Deze positieve ervaring vertaalt zich direct in een sterke bereidheid tot toekomstig gebruik. Deelnemers gaven met een zeer hoge score aan vaker VR te willen gebruiken voor testen of trainingen (Figuur



Figuur 5.1: Gemiddelde Likert-scores (schaal 1–5) per vraag voor de VR-screening: Q1: Denk je dat je deze soort test langdurig of vaker zou kunnen uitvoeren? Q2: Heb je tijdens of na de VR-ervaring last gehad van duizeligheid of misselijkheid? Q3: Hoe leuk vond je het om de test in VR uit te voeren? Q4: Vond je de instructies voorafgaand aan de VR-test duidelijk? Q5: Was de software gemakkelijk te gebruiken? Q6: Zou je graag vaker VR gebruiken tijdens testen of trainingen?

5.1 Q6) (M=4.75, SD=0.50) en toonden zich ook positief over het potentieel voor langduriger of herhaaldelijk gebruik van deze specifieke test (Figuur 5.1 Q1) (M=4.0, SD=1.15).

5.1.2 Algemene perceptie en ervaring

De overkoepelende toon van de feedback was uitgesproken positief en getuigde van een hoge mate van engagement. Deelnemers gebruikten consequent positieve en enthousiaste bewoordingen om hun ervaring te beschrijven, zoals “Heel plezant en leerrijk”, “Heel interessant. Je leert er iets van”, en een “goede, leuke ervaring”.

De interactie met de virtuele omgeving, en in het bijzonder de keuze voor hand-tracking als besturingsmethode, vormde een belangrijk thema. De feedback was unaniem positief over deze keuze. Hand-tracking werd omschreven als een duidelijke “meerwaarde” en als intuïtiever dan het alternatief van controllers. Een deelnemer verwoordde de kern van dit voordeel treffend door te anticiperen op de moeilijkheden die controllers met zich mee zouden brengen: “Ik denk dat besturen met mijn handen gemakkelijker is, ik zou in de war komen met de verschillende knoppen.”

5.1.3 Suggesties voor verbetering

Er was een duidelijke wens voor een rijkere en meer functionele virtuele ruimte. Een deelnemer gaf aan behoefte te hebben aan “Meer invulling in de ruimte” omdat deze “te leeg” aanvoelde. Een specifieke opmerking betrof de belichting: “[De] schaduwen zijn te donker waardoor het soms moeilijk te zien is”.

5.2 Resultatendashboard voor begeleiders

Dit hoofdstuk presenteert de resultaten van de evaluatiestudie van het VR-dashboard. Het doel van deze studie was het meten van de effectiviteit, de bruikbaarheid en de bijdrage aan objectiviteit van dit systeem als hulpmiddel voor het beoordelen van competenties.

De bevindingen in dit hoofdstuk zijn gebaseerd op een gemengde analyse van kwantitatieve data, verzameld via de scores en zekerheidsbeoordelingen van de deelnemers in elke fase, en kwalitatieve data, verkregen uit de post-test vragenlijst met Likert-schaalvragen en open vragen. Scores werden gegeven op een 5-puntsschaal, waarbij 1 staat voor ‘helemaal niet akkoord’ / ‘geen last’ en 5 voor ‘helemaal akkoord’ / ‘zeer veel last’.

5.2.1 De impact van informatie op de beoordeling

Een kerndoel van de evaluatie was te onderzoeken hoe de beoordelingen van de deelnemers evolueerden naarmate zij meer informatie via het dashboard kregen. De analyse focust op twee aspecten: de zekerheid waarmee de deelnemers hun oordeel geven en de uiteindelijke beoordelingsscores die zij toekenden.

Analyse van de zekerheid van de beoordelaars

De kwantitatieve data tonen een duidelijke toename in de gerapporteerde zekerheid van de beoordelaars bij elke nieuwe informatiefase. Tabel 5.1 illustreert deze progressie.

Tabel 5.1: Gemiddelde Zekerheid (schaal 1-5) per evaluatiefase.

Fase	Gemiddelde zekerheid	Standaardafwijking Zekerheid
Overview	2.68	0.41
Plots	4.13	0.57
Simulatie	4.77	0.42

De gemiddelde zekerheid van de deelnemers steeg significant van M=2.68 (SD=0.41) in de 'Overview'-fase naar M=4.13 (SD=0.57) na de toevoeging van de datavisualisaties (plots). Een t-toets bevestigt dat deze toename van 54.2% statistisch significant is ($t(22)$ = -7.15, $p < 0.01$). De introductie van de 3D-simulatie zorgde voor een verdere significante stijging naar M=4.77 (SD=0.42), een additionele toename van 15.3%. Ook deze stijging werd bevestigd als statistisch significant ($t(22)$ = -3.10, $p < 0.01$)

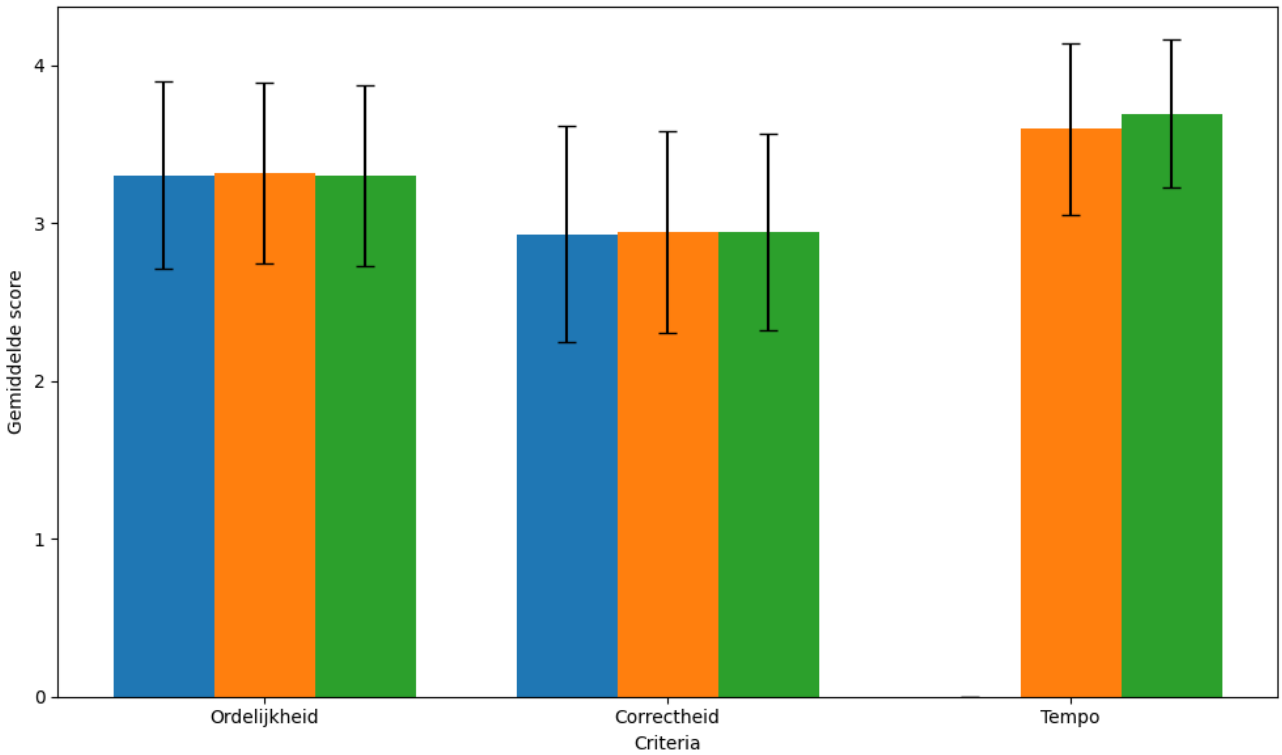
Een bijkomende belangrijke bevinding is de afname van de standaardafwijking van de zekerheidsscores in de laatste fase. Na de introductie van de plots nam de spreiding in zekerheid toe, maar na het bekijken van de simulatie daalde deze (van SD=0.57 naar SD=0.42). Dit suggereert dat de simulatie niet alleen de individuele zekerheid verhoogde, maar ook leidde tot een grotere consensus onder de beoordelaars. De lage standaardafwijking (SD = 0.41) in de eerste fase is het gevolg van het feit dat de competentie 'Tempo' pas in de tweede fase kon worden beoordeeld. De volledige overeenstemming hierover heeft geleid tot deze lage waarde, wat ook zichtbaar is in Figuur 5.2.

Deze kwantitatieve trend wordt krachtig ondersteund door de kwalitatieve feedback van zowel de expert als de andere deelnemers. Hoewel de grootste procentuele stijging in zekerheid plaatsvond na de introductie van de plots, gaven de respondenten overweldigend aan dat zij zich pas het meest zeker voelden na het zien van de 3D-simulatie. Op de post-test vragenlijst gaf de expert de stelling "Ik voelde me zekerder in mijn beoordeling naarmate ik meer informatie kreeg" de maximale score van 5. De plots boden gestructureerde data die hielpen bij het vormen van een eerste hypothese, en met de simulatie kon deze hypothese bevestigt of ontkracht worden. Respondent P11 verwoordde dit als volgt: "Na de grafieken dacht ik het meeste te weten, maar na de simulatie was ik het zekerste". De simulatie bood de context die de abstracte datapunten in de grafieken misten. Dit werd treffend samengevat door deelnemer P02: "In de simulatie zie je het proces en je moet het niet proberen te voorspellen uit de getallen en grafieken". De simulatie werd dus gezien als de meest complete en betrouwbare informatiebron, die de beoordelaars het vertrouwen gaf om een definitief oordeel te vellen.

Analyse van de beoordelingsscores

In tegenstelling tot sterke toenemende zekerheid, vertoonden de daadwerkelijke beoordelingsscores een ander patroon. Zoals te zien in Figuur 5.2, was er een aanzienlijke verandering in de gemiddelde score na de introductie van de plots in het criterium 'Tempo' (M=0 naar M=3.60). Deze sprong is voornamelijk te verklaren doordat het criterium 'Tempo' pas beoordeeld kon worden na het zien van de grafieken. De overgang van de 'Plots'-fase naar de 'Simulatie'-fase resulteerde in een slechts marginale verandering van de gemiddelde score.

Deze stabiliteit van de scores suggereert dat de extra informatie die de simulatie bood, voornamelijk diende ter bevestiging van het oordeel dat al was gevormd op basis van het overzicht en de plots, in plaats van dat het leidde tot een complete herziening. Dit wordt direct ondersteund



Figuur 5.2: Gemiddelde Beoordelingsscore (schaal 1-4) van overzicht (blauw), plots (oranje) en simulatie (groen) per criteria.

door de resultaten uit de post-test vragenlijst. De stelling "Mijn beoordeling veranderde zodra ik meer informatie van het systeem kreeg" ontving een neutrale gemiddelde score van M=3.0, wat aangeeft dat een verandering van oordeel niet de dominante ervaring was.

De kwalitatieve antwoorden van de respondenten verhelderen deze bevinding. Meerdere deelnemers gaven expliciet aan dat de extra informatie hun oordeel niet beïnvloedde, maar wel hun zekerheid. Typische antwoorden waren: "Deze hadden geen invloed op de score, maar wel op mijn zekerheid (P05)" en "Niet beïnvloed, enkel bevestiging gegeven (P04)". Een andere deelnemer stelde: "Mijn score zelf is niet vaak veranderd, maar ik werd er wel zekerder van (P12)". De expert gaf aan dat de extra informatie diende ter "Verduidelijken van informatie" en dat hij de data voornamelijk gebruikte ter "bevestiging" van zijn oordeel. De data veranderden zijn deskundige mening dus niet fundamenteel; ze leverden het objectieve bewijs dat nodig was om die mening te onderbouwen. De primaire functie van de rijkere data, en met name de simulatie, was dus niet zozeer het corrigeren van een foutief oordeel, maar het versterken en valideren van een reeds gevormd oordeel. Het dashboard fungeerde in de praktijk dus meer als een instrument voor het opbouwen van vertrouwen dan als een instrument voor het veranderen van beslissingen.

Een kwantitatieve blik op consensusvorming

De data onthult dat het dashboard niet enkel de individuele zekerheid van de beoordelaars verhoogt, maar ook een convergerend effect heeft op hun uiteindelijke scores. Dit duidt op een toename in consensus. Hoewel de totale standaarddeviatie van de scores een complex beeld geeft, toont een analyse per beoordelingscriterium een duidelijke trend naar een grotere overeenstemming naarmate meer informatie beschikbaar wordt gesteld.

Voor de competentie 'Correctheid' daalt de gemiddelde standaarddeviatie consistent van 0.69 in de 'Overview'-fase naar 0.64 na de introductie van de 'Plots', en verder naar 0.62 na het bekijken van de 'Simulatie'. Een vergelijkbare trend is zichtbaar bij het criterium 'Tempo', waar de standaarddeviatie afneemt van 0.54 naar 0.47 tussen de 'Plots'- en 'Simulatie'-fase. Deze afname in spreiding is een sterke kwantitatieve indicator dat het dashboard fungeert als een objectiverend instrument. Het biedt een gemeenschappelijk, data-gedreven referentiekader dat de initiële, mogelijk uiteenlopende, indrukken van beoordelaars naar een meer uniforme en gedeelde conclusie stuurt.

Dit sluit aan bij de perceptie van de deelnemers, die het systeem een zeer hoge score gaven op de stelling "Dit systeem kan bijdragen aan meer objectieve beoordelingen" (M=4.58, SD=0.67). Desondanks blijft een belangrijke nuance bestaan, zoals verwoord door een van de respondenten: "Veel objectiever dan enkel observeren. Je moet de data alsnog juist interpreteren wat tot subjectiviteit kan leiden" (P01).

5.2.2 Bruikbaarheid en perceptie van het dashboard

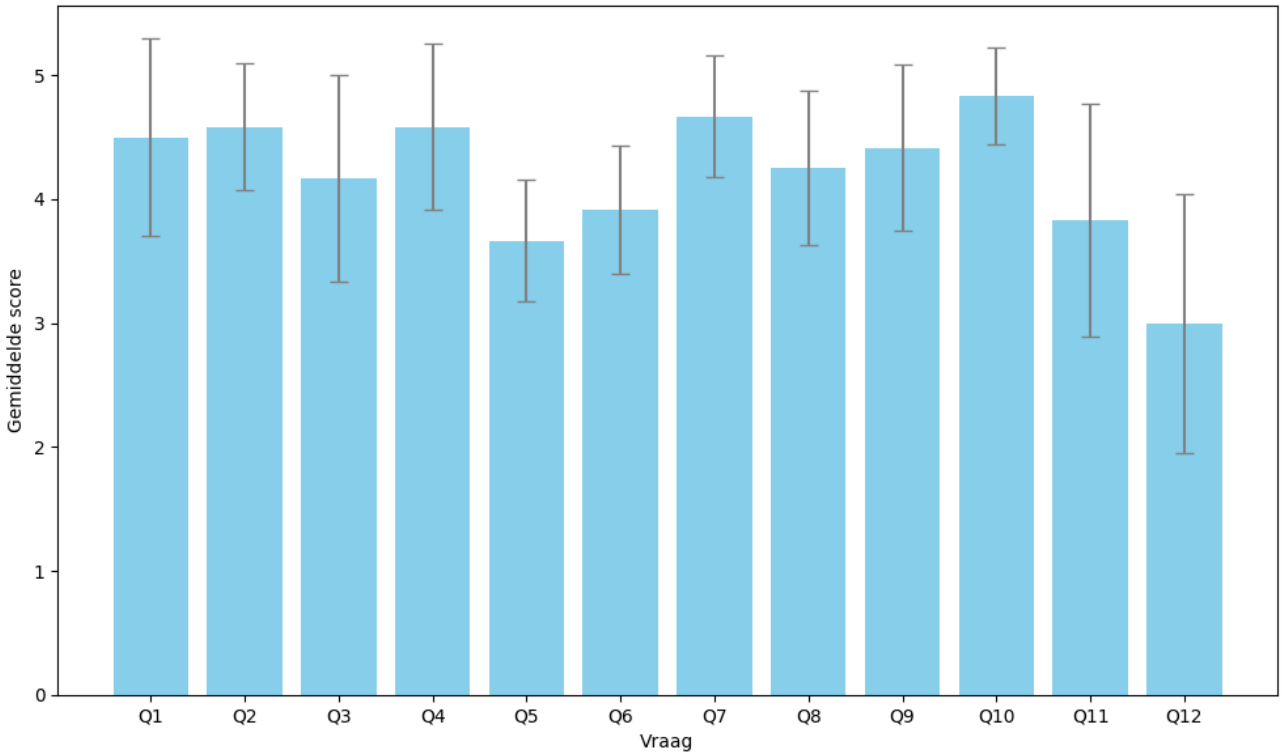
Naast de impact op het beoordelingsproces werd ook de algemene gebruikerservaring en de waarde van de verschillende dashboard-componenten geëvalueerd. Figuur 5.3 geeft een overzicht van de scores op de belangrijkste Likert-schaalvragen met betrekking tot de perceptie van het dashboard.

Meest nuttige componenten

Deelnemers waardeerden zowel de datavisualisaties (plots) als de 3D-simulatie zeer hoog, met gemiddelde scores van respectievelijk M=4.58 (SD=0.51) en M=4.50 (SD=0.90). De kwalitatieve feedback onthulde echter dat deze componenten niet als concurrerend werden gezien, maar als complementair, elk met een eigen functie binnen de beoordeling. Dit wordt ook ondersteund door de expert. Dit blijkt uit de maximale scores van 5 op de stellingen "De data plots hielpen mij bij het maken van een betere beoordeling" en "De 3D-simulatie gaf mij extra inzicht in de prestatie van de deelnemer".

Deelnemers gaven aan dat het 'Overzicht' vaak diende voor een snelle, eerste indruk. Vervolgens werden de 'Plots' gebruikt voor een efficiënte, data-gedreven analyse van trends, patronen en afwijkingen over de tijd. De 'Simulatie' werd ten slotte ingezet voor een diepgaande, contextuele analyse en voor de definitieve bevestiging van het oordeel. Deelnemer P12 beschreef deze workflow als volgt: "Het overzicht was vaak al voldoende om een score te bedenken. De grafieken en de simulatie gebruikte ik daarna om mijn gedachten te bevestigen". Dit wijst erop dat de kracht van het dashboard niet in een enkele component schuilt, maar in de combinatie ervan, die de beoordelaar in staat stelt om van een globaal overzicht naar een gedetailleerde analyse te gaan. Dit werd treffend samengevat door een deelnemer P10: "Het is de combinatie die duidelijkheid geeft, er is geen enkel onderdeel waarvan ik denk dat ik het niet nodig had".

De expert gaf aan dat de combinatie van plots en simulatie hem het meest zeker maakte. Hij gebruikte de plots om objectieve patronen en afwijkingen te detecteren: "Als ik statistieken zie, dan ga ik beter kunnen zien waar je aandacht of tijd aan gegeven hebt". Vervolgens wendde hij zich tot de simulatie voor een diepgaandere analyse en de definitieve bevestiging van zijn oordeel, waarbij hij opmerkte dat "Beeldmateriaal nodig [is] om te zien wat daar gebeurd is"



Figuur 5.3: Gemiddelde Likert-scores (schaal 1–5) per vraag voor het resultatendashboard: Q1: De 3D-simulatie bood extra inzicht in de prestaties van de deelnemer; Q2: De datavisualisaties ondersteunden het maken van een betere beoordeling; Q3: De systeemscore was duidelijk en goed onderbouwd; Q4: Het systeem kan bijdragen aan meer objectieve beoordelingen; Q5: Het systeem lijkt in staat zelfstandig correcte beoordelingen te maken; Q6: Het dashboard was gebruiksvriendelijk; Q7: Het systeem verschafte informatie die niet waarneembaar is tijdens een fysieke uitvoering; Q8: De beschikbare informatie was toereikend voor een objectieve beoordeling; Q9: Meer informatie verhoogde mijn beoordelingszekerheid; Q10: Ik zou het systeem inzetten als hulpmiddel bij screenings; Q11: Ik zou het systeem vertrouwen voor autonome beoordelingen; Q12: Mijn beoordeling veranderde na ontvangst van extra systeeminformatie.

RQ1: Gebruiksgemak en duidelijkheid

De algemene bruikbaarheid van het dashboard werd positief beoordeeld, met een score van M=3.92 (SD=0.51) op de stelling "Het dashboard was eenvoudig te gebruiken". Deze score, hoewel niet zeer laag, duidt erop dat er ruimte is voor verbetering. De kwalitatieve feedback verschafte concrete inzichten in de specifieke punten die als verwarrend of moeilijk werden ervaren.

Meerdere respondenten gaven aan dat de grafieken soms moeilijk te interpreteren waren zonder extra uitleg. Specifieke problemen waren het ontbreken van duidelijke labels en de complexiteit van bepaalde plots, zoals deelnemer P06 opmerkte: "De labels in de grafieken ontbraken, wat documentatie of tooltips zouden handig zijn, aangezien je veel moest uitleggen.". Ook de terminologie in het foutenoverzicht was niet voor iedereen onmiddellijk duidelijk, bijvoorbeeld het onderscheid tussen verschillende fouttypes.

De expert merkte op dat de verwoording in het overzicht voor verwarring kon zorgen en dat sommige grafieken baat zouden hebben bij betere labels. Een specifiek punt van verwarring was de kleurcodering, waarbij de betekenis niet altijd eenduidig was: "Rood kan zowel een foute doos als juist aanduiden, als juist als fout aanduiden", waarop hij concludeerde dat er een verschil gemaakt moet worden.

Een bijkomende suggestie van de expert was dat het systeem niet alleen moet focussen op fouten, maar ook correcte handelingen moet benadrukken om een gebalanceerd en motiverend beeld te geven. De expert merkte op dat het "Positieve [...] mee getoond [moet] kunnen worden".

De deelnemers formuleerden diverse constructieve suggesties voor verbetering. De meest genoemde suggestie betrof de besturing van de 3D-simulatie. Er was een sterke wens voor een meer intuïtieve navigatie, vergelijkbaar met moderne videospelers: "Doorspoelen in de simulatie door gebruik te maken van de balk (P03)" was een veelgevraagde functie. Een zinvolle aanbeveling was het creëren van een directe koppeling tussen de datapunten in de grafieken en de simulatie. Deelnemer P08 stelde voor: "Een link van de plot naar de simulatie wanneer je op een bolletje klikt, zo kan je vlug zien wat er daar aan het gebeuren is". Dergelijke functionaliteit zou de workflow, waarbij de plots worden gebruikt om anomalieën te detecteren en de simulatie om ze te onderzoeken, aanzienlijk versnellen.

RQ2: Bijdrage aan objectiviteit

De deelnemers en de expert waren zeer positief over het potentieel van het dashboard om de objectiviteit van beoordelingen te verhogen. De stelling "Dit systeem kan bijdragen aan meer objectieve beoordelingen" (Figuur 5.3) kreeg een score van M=4.58 (sd=0.67). Dit werd verder versterkt door de score van M=4.67 (SD=0.49) op de stelling "Het systeem gaf me informatie die ik niet zou kunnen waarnemen in een fysieke uitvoering", wat de unieke meerwaarde van de data-gedreven aanpak onderstreept.

De kwalitatieve antwoorden bieden echter een belangrijke nuance op dit positieve beeld. Hoewel de deelnemers de data die het systeem genereert als inherent objectief beschouwen, erkennen zij dat de interpretatie van die data een menselijke, en dus deels subjectieve, handeling blijft. Respondent P01 vatte dit dilemma samen: "Veel objectiever dan enkel observeren, Je moet de data alsnog juist interpreteren wat tot subjectiviteit kan leiden". Deelnemers gaven aan dat het systeem perfect laat zien wat er gebeurt, maar niet waarom. Deze behoefte aan context en het begrijpen van de intentie achter een actie kwam herhaaldelijk naar voren. Een citaat dat dit

illustreert is: “Je ziet niet altijd alles in de simulatie, de menselijke factor is ook een onderdeel, dus kunnen vragen aan de persoon waarom bepaalde dingen gebeuren zou een voordeel zijn (P09)”. Het systeem wordt dus gezien als een krachtig instrument dat objectieve bewijslast levert, maar de finale oordeelsvorming vereist nog steeds een menselijke interpretatielaag.

5.2.3 Vertrouwen in autonome beoordeling

De analyse van het vertrouwen in het systeem onthulde een van de meest significante bevindingen van deze studie: een duidelijk onderscheid tussen het vertrouwen in het systeem als hulpmiddel en het vertrouwen in het systeem als autonome beoordelaar. Er was een bijna unanieme consensus over de waarde van het dashboard als ondersteunend instrument. De stelling “Ik zou dit systeem gebruiken als hulpmiddel bij screenings” behaalde een uitzonderlijk hoge gemiddelde score van M=4.83 (SD=0.39). Dit staat in schril contrast met de scores met betrekking tot volledige autonomie. De stellingen “Ik zou het systeem vertrouwen om personen autonoom te beoordelen” en “Dit systeem lijkt me in staat om zelf correcte beoordelingen te kunnen maken” kregen aanzienlijk lagere en meer verdeelde scores van respectievelijk M=3.83 (SD=0.94) en M=3.67 (SD=0.49).

Om dit waargenomen contrast te valideren, werd het gemiddelde van de twee ‘autonomie’-stellingen vergeleken met de ‘hulpmiddel’-stelling. Een onafhankelijke t-toets toonde aan dat het vertrouwen in het systeem als hulpmiddel significant hoger was dan het vertrouwen in het systeem als autonome beoordelaar ($t(22) = 4.61$, $p < 0.01$).

De terughoudendheid ten aanzien van volledige automatisering lijkt niet primair voort te komen uit een wantrouwen in de accuraatheid van het algoritme, maar uit de wens van de beoordelaar om de controle te behouden en de conclusies van het systeem te kunnen verifiëren. Dit sluit aan bij de kernintentie van het project: niet het vervangen van mensen, maar het ondersteunen van beoordelaars in het maken van een betere en meer objectieve beoordeling.

De kwalitatieve data tonen aan dat beoordelaars de plots en de simulatie willen gebruiken om te begrijpen hoe het systeem tot zijn score komt. Een autonome score, gepresenteerd als een ‘black box’, ontneemt hen deze cruciale stap van validatie en begrip. Deze stap is belangrijk voor het opbouwen van hun eigen zekerheid. Het vertrouwen in het systeem is dus sterk gekoppeld aan de mogelijkheid om de resultaten te kunnen verifiëren. Respondent P03 verwoordde deze behoefte aan een ‘human-in-the-loop’ perfect: “Ik zie het voordeel van automatische score, maar ik zou de rest alsnog altijd bekijken om er zeker van te zijn dat deze niks over het hoofd heeft gezien”. Gebruikers vertrouwen het systeem als een transparante en verifieerbare assistent, maar zijn huiverig om de eindverantwoordelijkheid voor de beoordeling volledig over te dragen.

5.2.4 Analyse van specifieke gedragsprofielen

Terwijl de geaggregeerde resultaten de algemene trends en de effectiviteit van het dashboard aantonen, onthult een analyse van de afzonderlijke sessies (A-F) de ware diagnostische kracht van het systeem. Door in te zoomen op specifieke gedragsprofielen wordt duidelijk hoe het dashboard begeleiders in staat stelt om genuanceerde, context-specifieke inzichten te verkrijgen die essentieel zijn voor het geven van gerichte, persoonlijke feedback. De volgende subsecties bespreken een aantal van de meest illustratieve gevallen.

Vaststellen van een foutloze baseline

Sessie A, gedefinieerd als het profiel 'Alles juist', dient als de 'gouden standaard' voor een correcte uitvoering. De data uit het dashboard bevestigt dit profiel feilloos. De beoordelingen van de deelnemers weerspiegelen een hoge mate van competentie: de gemiddelde scores voor 'Correctheid' (M=3.83, SD=0.58) en 'Orde' (M=3.75, SD=0.45) zijn consistent hoog. De beoordeling van de expert is nog stelliger, met een perfecte score van 4 op 4 voor zowel 'Correctheid' als 'Orde' en een maximale zekerheid van 5 op 5 na het zien van de plots.

Vinden van initieel onzichtbare inconsistentie

Het profiel 'Afgeleid' is gedefinieerd als een persoon die correct werk aflevert, maar met een inconsistent en vertragend werktempo. Dit scenario illustreert perfect hoe het dashboard procesfouten kan blootleggen die in het eindresultaat volledig onzichtbaar zijn. De data toont dit contrast scherp aan: de score voor 'Correctheid' is hoog (M=3.92, SD=0.29), wat duidt op een kwalitatief goed eindproduct. De score voor 'Tempo' is met M=2.92 (SD=1.00) echter lager dan bij de meeste andere scenario's.

Het begrijpen van beoordelingsfouten

Een vergelijking tussen Sessie B en Sessie C toont de diagnostische precisie van het dashboard aan. In sessie B worden fouten gemaakt doordat incorrecte items toch als correct worden gelabeld (false negatives). De simulatie zou een gebruiker tonen die snel en oppervlakkig inspecteert, zonder de afwijkingen op te merken. In sessie C worden fouten gemaakt doordat correcte items onterecht als fout worden gemarkeerd (false positives). De simulatie toont een heel ander gedrag: een gebruiker die lang en aarzelend naar correct item kijkt en het uit onzekerheid of een te strikte interpretatie van de regels afkeurt.

Beide profielen leiden tot een lage score op 'Correctheid' (M=2.0 voor Sessie B en M=2 voor Sessie C), maar de onderliggende oorzaak van de fouten is fundamenteel verschillend. De data van de expertbeoordeling voor Sessie C bevat een van de meest sprekende voorbeelden van de meerwaarde van de simulatie. Terwijl in de meeste sessies de simulatie het oordeel van de expert bevestigde, leidde het in dit geval tot een herziening. De expert gaf initieel een zeer lage score van 1 op 4 voor 'Correctheid', waarschijnlijk gebaseerd op het hoge aantal afgekeurde items. Na het zien van de simulatie, die de onderliggende onzekerheid van de gebruiker blootlegde, verhoogde de expert zijn score naar 2 op 4. Dit toont aan dat de simulatie de context bood die nodig was om het gedrag correct te interpreteren, wat leidde tot een genuanceerder oordeel.

Hoofdstuk 6

Besluit

Dit hoofdstuk heeft tot doel de resultaten, zoals gepresenteerd in hoofdstuk 5, te interpreteren en in een breder kader te plaatsen. Waar het vorige hoofdstuk de bevindingen presenteerde, zal deze discussie de betekenis ervan duiden, verbanden leggen tussen de verschillende onderzoekscomponenten, en de theoretische en praktische implicaties van de studie bespreken. De structuur van dit hoofdstuk volgt een logische opbouw. Eerst wordt de validiteit van de VR-screening als meetinstrument geanalyseerd, aangezien dit het fundament vormt voor de betrouwbaarheid van de verzamelde data. Vervolgens wordt een diepgaande analyse gewijd aan de impact van het resultatendashboard op de beoordelaars, met een focus op de psychologische processen van oordeelsvorming en vertrouwen. Ten slotte worden de implicaties van de studie besproken, samen met de erkende beperkingen en concrete aanbevelingen voor zowel de doorontwikkeling van het systeem als voor toekomstig onderzoek.

6.1 Discussie

6.1.1 Gebruikerservaring als fundament voor de validiteit van competentiemetingen

De resultaten van de gebruikersevaluatie van de VR-screening tonen een positieve ervaring. Deze bevinding vormt de methodologische hoeksteen die de validiteit van de verzamelde prestatiedata garandeert. Zonder de hoge mate van acceptatie, gebruiksgemak en fysiek comfort zou de data een vertekend beeld kunnen geven van de werkelijke competenties van de deelnemers, waardoor de fundamenten van de gehele studie ondermijnd zouden worden.

De kwantitatieve data onderbouwen deze conclusie met grote eenduidigheid. De helderheid van de instructies werd als perfect beoordeeld, wat essentieel is om te verzekeren dat deelnemers de taak volledig begrepen alvorens te starten. Dit elimineert de storende variabele dat een slechte prestatie te wijten zou zijn aan onbegrip in plaats van een gebrek aan competentie. Het gebruiksgemak van de software zelf werd eveneens zeer hoog beoordeeld, wat wijst op een intuïtief ontwerp dat geen significante cognitieve belasting vormde voor de gebruikers. De vraag naar duizeligheid of misselijkheid leverde een perfecte score op, wat aangeeft dat geen enkele deelnemer fysiek ongemak had. Dit is cruciaal, aangezien fysiek ongemak de prestaties direct en negatief beïnvloedt, los van de te meten competenties. De positieve respons, met een hoge score voor plezier, en de sterke bereidheid tot toekomstig gebruik bevestigen de algehele succesvolle

implementatie.

Deze positieve gebruikerservaring kan worden beschouwd als een indicator van de ecologische validiteit van de meting. Een meting is pas valide als deze meet wat ze beoogt te meten. In een technologische context zoals VR betekent dit dat het systeem zelf zo naadloos en frictieloos mogelijk moet zijn. Indien een systeem fysiek ongemak of cognitieve frictie (onduidelijke instructies, complexe besturing) veroorzaakt, meet men niet langer primair de competentie van de gebruiker voor de taak, maar diens vermogen om met een gebrekkig technologisch instrument om te gaan.

Een diepere analyse van de kwalitatieve feedback onthult dat de keuze voor hand-tracking een cruciale ontwerpbeslissing was die de technologie toegankelijk maakte voor de specifieke doelgroep van maatwerkers. De feedback was positief over hand-tracking, omschreven als een "meerwaarde" en intuïtiever dan controllers. Voor een doelgroep die mogelijk minder ervaring heeft met complexe interfaces zoals gamecontrollers, verlaagt de keuze voor de meest natuurlijke interactiemethode, de eigen handen, de technologische drempel significant. Hierdoor kan de gebruiker zich volledig concentreren op de uit te voeren taak in plaats van op de bediening van het systeem.

6.1.2 De rol van data in oordeelsvorming

De evaluatie van het resultatendashboard onthult een van de meest opvallende bevindingen van deze studie: het psychologische effect van data op het beoordelingsproces. De resultaten wijzen erop dat de toegevoegde data niet in de eerste plaats dient om initiële oordelen te corrigeren, maar vooral werkt als een krachtig middel om de zekerheid over een reeds gevormd oordeel aanzienlijk te versterken en te bevestigen.

De kwantitatieve data illustreren een duidelijk patroon. De gerapporteerde zekerheid van de beoordelaars kende een statistisch significante toename in elke fase van het proces. Dit staat in schril contrast met de stabiliteit van de daadwerkelijke beoordelingsscores. De expert verwoordde dit door te stellen dat de data voornamelijk diende ter "bevestiging" van zijn oordeel. De data veranderden zijn deskundige mening niet fundamenteel, maar leverden het objectieve bewijs dat nodig was om die mening te onderbouwen.

Deze bevinding roept echter ook een belangrijke discussie op over het fenomeen ´confirmation bias', de menselijke neiging om informatie te zoeken, te interpreteren en te onthouden op een manier die de eigen, reeds bestaande overtuigingen bevestigt. Het feit dat beoordelaars hun oordeel zelden aanpasten na het ontvangen van meer data kan betekenen dat hun initiële oordeel, gevormd in de 'Overview'-fase, doorgaans correct was. Het kan echter ook inhouden dat zij de plots en de simulatie selectief interpreteerden om hun eerste indruk te bevestigen. De studie kan tussen deze twee verklaringen geen definitief uitsluitsel geven. Een belangrijke nuance hierbij is de casus van de expert die zijn oordeel voor Sessie C wél aanpaste na het zien van de simulatie, van een score 1 naar 2 op 4 voor 'Correctheid'. Dit toont aan dat het systeem wel degelijk kan leiden tot een herziening van het oordeel, maar dat dit mogelijk een hoger niveau van expertise of een meer open, kritische houding van de beoordelaar vereist. Dit heeft belangrijke implicaties voor de implementatie van dergelijke systemen in de praktijk. Begeleiders moeten getraind worden om het dashboard niet louter als een bevestigingsinstrument te gebruiken, maar ook als een instrument om hun eigen initiële oordelen kritisch tegen het licht te houden en actief op zoek te gaan naar data die hun hypotheses mogelijk ontkrachten.

6.1.3 Samenwerking tussen overzicht, datavisualisaties en simulatie

De effectiviteit van het resultatendashboard schuilt niet in één superieure component, maar in de synergetische combinatie van het overzicht, de datavisualisaties (plots) en de 3D-simulatie. De resultaten tonen aan dat deze componenten niet als concurrerend, maar als complementair werden ervaren, waarbij elke component een unieke rol vervult binnen een natuurlijke, gelaagde analytische workflow. Ook de expert bevestigde deze gelaagde aanpak: hij gebruikte de plots om objectieve patronen en afwijkingen te detecteren en gebruikte vervolgens de simulatie voor de contextuele interpretatie.

6.1.4 Vertrouwen door transparantie

De studie onthult een cruciale opmerking met betrekking tot het vertrouwen in het systeem. Terwijl het vertrouwen in het dashboard als ondersteunend hulpmiddel positief, heerst er terughoudendheid om het systeem als een volledig autonome beoordelaar te accepteren. Dit verschil lijkt niet voort te komen uit een wantrouwen in de accuraatheid van de data, maar uit een fundamentele menselijke behoefte aan transparantie, verifieerbaarheid en controle in besluitvormingsprocessen met een hoge mate van verantwoordelijkheid.

Deze bevinding sluit aan bij de kernprincipes van 'Explainable AI' (XAI), die stellen dat voor de acceptatie van AI-systemen in kritische domeinen, de systemen interpreteerbaar en transparant moeten zijn. De resultaten bevestigen de ontwerpfilosofie van dit project: de mens niet vervangen, maar ondersteunen.

Verder wordt de notie van objectiviteit door de deelnemers genuanceerd. Hoewel het systeem hoog scoort op de stelling dat het kan "bijdragen aan meer objectieve beoordelingen", erkennen de beoordelaars dat de finale oordeelsvorming een menselijke handeling blijft. Een deelnemer merkte op dat het systeem perfect laat zien wat er gebeurt, maar niet waarom, en dat de mogelijkheid om vragen te stellen aan de persoon een belangrijk, ontbrekend onderdeel is. Het dashboard faciliteert een objectiever proces en leidt tot een grotere consensus onder beoordelaars, zoals blijkt uit de afnemende standaarddeviaties voor de scores 'Correctheid' en 'Tempo' naarmate meer informatie beschikbaar wordt. De eindverantwoordelijkheid voor een eerlijk en contextueel juist oordeel blijft echter bij de menselijke expert liggen.

6.2 Praktische en theoretische implicaties

Op praktisch vlak heeft het ontwikkelde systeem het potentieel om de screening en beoordeling van competenties binnen maatwerkbedrijven en vergelijkbare organisaties fundamenteel te verbeteren. Het biedt een gestandaardiseerde, data-gedreven methode die objectiever, efficiënter en diagnostisch preciezer is dan traditionele observatiemethoden. Het stelt begeleiders in staat om niet alleen een score toe te kennen, maar ook om de onderliggende oorzaken van prestaties te begrijpen. Dit maakt het mogelijk om gerichtere, op bewijs gebaseerde en gepersonaliseerde feedback en coaching te geven, wat uiteindelijk kan leiden tot effectievere trajecten voor de doelgroep.

Op theoretisch vlak levert de studie bijdragen aan de vakgebieden Human-Computer Interaction (HCI) en data-ondersteunde besluitvorming. Het toont aan dat de waarde van dergelijke systemen niet alleen ligt in het veranderen van meningen, maar ook in het versterken van zekerheid

en vertrouwen. Ten tweede onderstreept de studie het cruciale belang van transparantie en verifieerbaarheid (het 'human-in-the-loop' principe) voor het opbouwen van vertrouwen in en de acceptatie van geautomatiseerde beoordelingssystemen. Het toont aan dat gebruikers controle en inzicht willen behouden, en dat 'black box'-systemen op weerstand stuiten in contexten met een hoge menselijke verantwoordelijkheid.

6.3 Beperkingen van de studie

De evaluatie van de VR-applicatie werd uitgevoerd met een kleine, specifieke groep van vier maatwerkers. Hoewel de resultaten zeer eenduidig en positief waren, is de generaliseerbaarheid naar een bredere populatie beperkt. Een grotere en meer diverse steekproef is nodig om deze bevindingen te valideren.

De studie is uitgevoerd binnen de specifieke context van één organisatie (Bewel) en één type taak. De bevindingen over de effectiviteit van het dashboard en de VR-tool zijn mogelijk niet direct overdraagbaar naar andere organisaties, doelgroepen of taakcontexten.

Het hoge engagement en de positieve ervaringen met de VR-tool kunnen deels beïnvloed zijn door de nieuwigheid. Het is mogelijk dat de motivatie en het enthousiasme afnemen naarmate de technologie meer gemeengoed wordt. Longitudinaal onderzoek is nodig om te bepalen of de positieve houding standhoudt bij herhaaldelijk en langdurig gebruik.

De studie meet de verandering in het oordeel en de zekerheid van de beoordelaars. Het kan echter niet definitief vaststellen of het uiteindelijke, door data ondersteunde oordeel ook daadwerkelijk 'correcter' is dan het initiële oordeel. Het definiëren van een objectieve 'ground truth' is hiervoor nodig.

6.4 Aanbevelingen voor vervolgontwikkeling en onderzoek

Op basis van de gedetailleerde feedback van zowel de deelnemers als de expert kunnen meerdere concrete aanbevelingen worden gedaan voor de verdere doorontwikkeling van het systeem.

Allereerst verdient het aanbeveling om de virtuele omgeving te verrijken met meer contextuele objecten die relevant zijn voor de werkplek. Daarnaast kan het verbeteren van de belichting en schaduw-rendering bijdragen aan een hogere zichtbaarheid en een realistischer weergave.

Verder zou het implementeren van een interactieve tijdlijn in de 3D-simulatie, vergelijkbaar met die in moderne videospelers, de gebruiker in staat stellen om snel en intuïtief door de simulatie te navigeren. Ook het toevoegen van duidelijke as-labels, een legende en contextuele tooltips aan alle datavisualisaties kan de interpreteerbaarheid aanzienlijk verhogen.

Om de consistentie en duidelijkheid te bevorderen, is het ontwikkelen van een eenduidige kleurcoderingsstrategie voor het gehele dashboard wenselijk, waarbij een duidelijke legende wordt toegevoegd die de betekenis van elke kleur uitlegt. Tot slot zou het introduceren van een click-to-jump-functionaliteit, die datapunten in de grafieken direct koppelt aan het corresponderende tijdstip in de 3D-simulatie, de efficiëntie en gebruiksvriendelijkheid van het systeem verder verbeteren.

Naast de concrete systeemverbeteringen levert deze studie ook aanbevelingen op voor toekomstig onderzoek.

Ten eerste is het aan te bevelen om longitudinale studies uit te voeren waarbij het systeem gedurende een langere periode, bijvoorbeeld enkele maanden, wordt gebruikt. Dit maakt het mogelijk om het zogenaamde novelty effect uit te sluiten en te onderzoeken hoe interactie, vertrouwen en de effectiviteit van de feedback zich ontwikkelen bij routinematig gebruik.

Daarnaast kan een gecontroleerde, vergelijkende studie waardevolle inzichten opleveren door de effectiviteit, efficiëntie en objectiviteit van beoordelingen via het dashboard direct te vergelijken met traditionele observatiemethoden. Op deze manier kan de impact op de juistheid van beoordelingen nauwkeurig worden gekwantificeerd.

Verder verdient het aanbeveling om in te gaan op de waarom-vraag achter het geobserveerde gedrag. Waar de huidige opzet zich richt op het meten van wat iemand doet, zou toekomstig onderzoek ook de intenties achter dit gedrag moeten blootleggen. Dit kan bijvoorbeeld door het integreren van think-aloud-protocollen tijdens de VR-test of door gestructureerde interviews direct na de sessie. Het gebruik van aanvullende meetmethoden zoals eye-tracking, EEG of GSR kan hierbij eveneens waardevolle inzichten bieden.

6.5 Slotopmerkingen

Deze masterproef startte met de ambitie om de traditionele, observatiegestuurde competentiebeoordeling binnen Bewel te verrijken door de inzet van virtual reality. Het onderzoek heeft aangetoond dat een VR-gebaseerd screeningssysteem niet alleen technisch en functioneel haalbaar is, maar ook een aanzienlijke meerwaarde biedt. Door het proces van dataverzameling te automatiseren en te objectiveren, wordt de basis gelegd voor een eerlijker en consistenter evaluatieproces.

De ontwikkelde tool is een instrument dat de menselijke expertise van de begeleider centraal stelt en versterkt. De studie toont aan dat begeleiders het systeem niet zien als een autonome vervanger, maar als een krachtig hulpmiddel dat hen in staat stelt beter onderbouwde, transparantere en uiteindelijk effectievere feedback te geven.

De positieve ontvangst door de maatwerkers zelf onderstreept het potentieel van VR als een inclusieve en motiverende technologie. De keuze voor intuïtieve interactiemethoden zoals handtracking bleek cruciaal om de technologische drempel te verlagen en een valide meting te garanderen. Dit project bewijst dat technologie, mits doordacht en mensgericht ontworpen, een sleutelrol kan spelen in het creëren van gelijkwaardige kansen en het ondersteunen van de professionele groei van elke medewerker, ongeacht diens achtergrond of vaardigheidsniveau.

Literatuurlijst

[1] N. Lander, D. Nahavandi, S. Mohamed, I. Essiet, and L. M. Barnett, "Bringing objectivity to motor skill assessment in children," vol. 38, no. 13, pp. 1539–1549. Publisher: Routledge _eprint: https://doi.org/10.1080/02640414.2020.1747743.

[2] C. C. T. Clark, M. C. Bisi, M. J. Duncan, and R. Stagni, "Technology-based methods for the assessment of fine and gross motor skill in children: A systematic overview of available solutions and future steps for effective in-field use," vol. 39, no. 11, pp. 1236–1276. Publisher: Routledge _eprint: https://doi.org/10.1080/02640414.2020.1864984.

[3] K. Lam, J. Chen, Z. Wang, F. M. Iqbal, A. Darzi, B. Lo, S. Purkayastha, and J. M. Kinross, "Machine learning for technical skill assessment in surgery: a systematic review," vol. 5, no. 1, p. 24. Publisher: Nature Publishing Group.

[4] P. V. González-Erena, S. Fernández-Guinea, and P. Kourtesis, "Cognitive assessment and training in extended reality: Multimodal systems, clinical utility, and current challenges," vol. 5, no. 1, p. 8. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[5] L. M. Daling and S. J. Schlittmeier, "Effects of augmented reality-, virtual reality-, and mixed reality–based training on objective performance measures and subjective evaluations in manual assembly tasks: A scoping review," vol. 66, no. 2, pp. 589–626. Publisher: SAGE Publications Inc.

[6] D. Scorgie, D. Paes, Z. Feng, and F. Parisi, "Virtual reality for safety training: A systematic literature review and meta-analysis,"

[7] C. Ladha, N. Y. Hammerla, P. Olivier, and T. Plötz, "ClimbAX: skill assessment for climbing enthusiasts," in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, UbiComp '13, pp. 235–244, Association for Computing Machinery.

[8] A. N. Neher, F. Bühlmann, M. Müller, C. Berendonk, T. C. Sauter, and T. Birrenbach, "Virtual reality for assessment in undergraduate nursing and medical education – a systematic review," vol. 25, p. 292.

[9] V. Li, I. Siniosoglou, P. Sarigiannidis, and V. Argyriou, "Enhancing manufacturing training through VR simulations."

[10] J. Chan, D. J. Pangal, T. Cardinal, G. Kugener, Y. Zhu, A. Roshannai, N. Markarian, A. Sinha, A. Anandkumar, A. Hung, G. Zada, and D. A. Donoho, "A systematic review of virtual reality for the assessment of technical skills in neurosurgery," Section: Neurosurgical Focus.

[11] I. S. Fitton, E. Dark, M. M. Oliveira da Silva, J. Dalton, M. J. Proulx, C. Clarke, and C. Lutteroth, "Watch this! observational learning in VR promotes better far transfer than active learning for a fine psychomotor task," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, pp. 1–19, Association for Computing Machinery.

[12] P. Strojny and N. Dużmańska-Misiarczyk, "Measuring the effectiveness of virtual training: A systematic review," vol. 2, p. 100006.

[13] Y. J. Yi, N. Heidari Matin, D. Brannan, M. Johnson, and A. Nguyen, "Design considerations for virtual reality intervention for people with intellectual and developmental disabilities: A systematic review," vol. 17, no. 4, pp. 212–241.

[14] D. R. Sanchez, E. Weiner, and A. Van Zelderen, "Virtual reality assessments (VRAs): Exploring the reliability and validity of evaluations in VR," vol. 30, no. 1, pp. 103–125. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ijsa.12369.

[15] M. W. Schmidt, K. F. Köppinger, C. Fan, K. F. Kowalewski, L. P. Schmidt, J. Vey, T. Proctor, P. Probst, V. V. Bintintan, B. P. Müller-Stich, and F. Nickel, "Virtual reality simulation in robot-assisted surgery: meta-analysis of skill transfer and predictability of skill," vol. 5, no. 2, p. zraa066.

[16] N. Gavish, T. Gutierrez, and S. Webel, "Design guidelines for the development of virtual reality and augmented reality training systems for maintenance and assembly tasks,"

[17] T. H. Turner, A. Atkins, and R. S. E. Keefe, "Virtual reality functional capacity assessment tool (VRFCAT-SL) in parkinson's disease," vol. 11, no. 4, pp. 1917–1925.

[18] W. Wong, N. Tan, L. Jie, and J. Allen, "Comparison of time taken to assess cognitive function using a fully immersive and automated virtual reality system vs. the montreal cognitive assessment,"

[19] K. R. Vijayanagar, "I, p, and b-frames - differences and use cases made easy - OTTVerse." Section: Compression.