



Fachhochschule Nordwestschweiz
Hochschule für Angewandte Psychologie

Play & Predict: Entwicklung und Untersuchung eines Verhaltensankersystems zur Beurteilung überfachlicher Fähigkeiten im Kontext eines kooperativen Videospiele

BACHELOR THESIS

Mai 2025

Autor

Luca Streich

Betreuungsperson

Dr. Serge Petralito

Praxispartner

Netzwerk Grenchen, LevelHub

Cristina Gerber

Abstract

Hintergrund: In einer zunehmend vernetzten und kollaborativen Arbeitswelt gewinnen überfachliche Fähigkeiten wie Entscheidungsfindung, Teamarbeit und Führung an Bedeutung. Die sozialunternehmerische Initiative LevelHub nutzt kooperative Videospiele zur Beobachtung und Förderung solcher Kompetenzen. Zur Erfassung fehlten jedoch bisher standardisierte Instrumente. Ziel dieser Studie war die Entwicklung und Evaluation eines verhaltensbasierten Bewertungssystems (VAS-KOOP) zur Einschätzung überfachlicher Fähigkeiten im Kontext eines kooperativen Spiels.

Methode: Basierend auf dem Rahmenmodell Non-Technical Skills for Surgeons (NOTSS) wurden mittels Critical Incident Technique (CIT) spezifische Verhaltensindikatoren für drei Hauptkategorien (Entscheidungsfindung, Kommunikation & Teamarbeit und Führung) definiert. In einer Untersuchung mit 30 Proband:innen erfolgte eine strukturierte videogestützte Fremdeinschätzung durch zwei unabhängige Bewertende sowie eine ergänzende Selbsteinschätzung. Evaluationskriterien waren die interne Konsistenz, Interrater-Reliabilität und Konstruktvalidität des Instruments.

Resultate: Die Analyse der Items zeigt eine gute interne Konsistenz. Die Fremdeinschätzungen zweier unabhängiger Bewertenden erwiesen sich als konsistent, während die Zusammenhänge mit den Selbsteinschätzungen gering ausfielen.

Schlussfolgerung: Das VAS-KOOP kann als strukturiertes und reliables Instrument zur Beobachtung überfachlicher Kompetenzen im Kontext eines kooperativen Spiels eingesetzt werden. Es eignet sich für Anwendungen in der Teamentwicklung, Personalbeurteilung und zur Kompetenzentwicklung.

Keywords: Verhaltensankersystem, überfachliche Fähigkeiten, kooperative Videospiele, game-based Assessments

Anzahl Zeichen inkl. Leerzeichen (ohne Anhang): 107'890

Inhaltsverzeichnis

1	Einleitung	5
1.1	Ausgangslage	5
1.2	Problemstellung	6
1.3	Zielsetzung und Fragestellung	6
1.4	Aufbau der Arbeit	7
2	Theoretische Grundlagen	8
2.1	LevelHub.....	8
2.1.1	<i>Vorgehensweise</i>	8
2.1.2	<i>Warum Videospiele?</i>	10
2.1.3	<i>Game-based Assessment</i>	10
2.2	Verhaltensankersysteme	11
2.2.1	<i>Definition</i>	11
2.2.2	<i>Evidenzbasis und Anwendungsbereiche</i>	11
2.2.3	<i>Non-Technical Skills for Surgeons</i>	12
3	Methode	15
3.1	Entwicklung des VAS-KOOP	15
3.1.1	<i>Recherche zu passenden Modellen und Fähigkeiten</i>	15
3.1.2	<i>Modifikation eines etablierten Instruments</i>	16
3.1.3	<i>Verhaltensankersystem VAS-KOOP</i>	19
3.2	Datenerhebung	20
3.2.1	<i>Stichprobe</i>	20
3.2.2	<i>Ablauf der Datenerhebung</i>	21
3.2.3	<i>Videogestützte Beobachtungsanalyse</i>	23
3.3	Datenauswertung	25
4	Ergebnisse	27
4.1	Deskriptiv Analyse.....	27
4.2	Interne Konsistenz	29
4.3	Interrater-Reliabilität.....	30
4.4	Vergleich Fremd- und Selbsteinschätzung	31
4.5	Zusammenfassung der Ergebnisse	31
5	Diskussion	32
5.1	Zusammenfassung der Arbeit.....	32
5.2	Interpretation der Ergebnisse	33
5.3	Limitationen der Untersuchung.....	37
6	Schlussfolgerung	39
6.1	Implikationen für LevelHub	39
6.2	Implikationen für Teamentwicklung und Personalbeurteilung.....	39
6.3	Perspektiven für zukünftige Forschung.....	40
6.4	Fazit.....	41
7	Literaturverzeichnis	42
8	Abbildungsverzeichnis	48
9	Tabellenverzeichnis	48
10	Hilfsmittelverzeichnis mit Verwendungszweck	49

Abkürzungsverzeichnis

BARS	Behaviorally Anchored Rating Scale
CIT	Critical Incident Technique
GBA	Game-based Assessment
GBL	Game-based Learning
NOTSS	Non-Technical Skills for Surgeons
OECD	Organisation for Economic Co-operation and Development
VAS-KOOP	Verhaltensankersystems im Kontext eines kooperativen Videospiele

1 Einleitung

1.1 Ausgangslage

Was passiert eigentlich, wenn Menschen gemeinsam ein Videospiel spielen – und woran lässt sich erkennen, wie gut sie zusammenarbeiten? Solche Spielsituationen werfen nicht nur spannende Fragen zur Teamdynamik auf, sondern berühren auch zentrale Anforderungen an das Arbeiten der Zukunft. In einer zunehmend komplexen und dynamischen Arbeitswelt gewinnen überfachliche Fähigkeiten wie Kommunikation, Kollaboration, Entscheidungsfindung oder Problemlösen stark an Bedeutung. Studien zeigen, dass solche Kompetenzen nicht nur für die berufliche Leistungsfähigkeit zentral sind, sondern auch die persönliche Entwicklung und gesellschaftliche Teilhabe fördern (Griffin & Care, 2015; Schulz, 2008). Internationale Organisationen wie die OECD (2023) und das World Economic Forum (2025) betonen die Notwendigkeit, diese Schlüsselkompetenzen gezielt zu fördern und systematisch zu bewerten, um Menschen auf die Anforderungen des digitalen und kollaborativen Arbeitens vorzubereiten. Traditionelle Bildungsansätze greifen dabei oft zu kurz – es braucht innovative Formate, die solche Fähigkeiten praxisnah entwickeln und sichtbar machen.

LevelHub setzt an der Förderung dieser zukunftsrelevanten Kompetenzen an. Sie nutzen Videospiele in Workshops, um Fähigkeiten wie Kooperation, Entscheidungsfindung oder Führungsverhalten spielerisch sichtbar zu machen. Spezifische ausgewählte Spiele erfordern beobachtbare kooperative Problemlöseprozesse, die im Anschluss gemeinsam reflektiert und mit entsprechenden Lerninhalten verknüpft werden. Die spielerische Lernarena fördert schnelle Erfolgserlebnisse und stärkt so die Offenheit für anschließende Selbstreflexionen. Die LevelHub-Methodik hat sich bereits im Bereich der Arbeitsintegration als wirkungsvoll erwiesen, indem persönliche Stärken und Potenziale bewusst gemacht wurden. Auch Studien belegen positive Effekte von Videospiele auf die Entwicklung berufsrelevanter Kompetenzen (Beigi, Ayoozadeh & Shirmohammadi, 2025; Connolly, Boyle, MacArthur, Hainey & Boyle, 2012; Granic, Lobel & Engels, 2014).

Nach diversen Pilotversuchen und gesammelten Erfahrungen sieht LevelHub weiteres Potenzial in der Vorgehensweise. Neue Einsatzgebiete sollen erschlossen werden, beispielsweise im Bereich der Personalbeurteilung oder Teamentwicklung. Für eine solche Ausweitung ist jedoch eine Standardisierung und Skalierung der Methodik erforderlich, um die Vergleichbarkeit und Qualität der Beobachtungen langfristig zu sichern.

Gaming zählt heute zu den beliebtesten Freizeitaktivitäten und wird zunehmend in der Kompetenzdiagnostik genutzt. Game-based Assessments (GBA) nutzen Videospiele gezielt zur Erfassung von Kompetenzen, Fähigkeiten oder Wissen und gelten als motivierende Alternative zu klassischen Verfahren (Gomez, Ruipérez-Valiente & Clemente, 2023). In der Forschung konzentrieren sich viele Studien auf die Messung sozialer Kompetenzen mithilfe spielerischer Elemente (Bhatia & Ryan, 2018; Chamorro-Premuzic, Winsborough, Sherman & Hogan, 2016;

Landers & Sanchez, 2022). Erste Erkenntnisse bestätigen das Potenzial von GBAs für die Erfassung berufsrelevanter Fähigkeiten.

Dennoch besteht eine Forschungslücke im Hinblick auf die Bewertung überfachlicher Fähigkeiten in kooperativen Spielsituationen (Hussein, Ow, Elaish & Jensen, 2022). Es fehlen bislang standardisierte Verfahren, um das beobachtbare Verhalten von Menschen, die gemeinsam vor einem Bildschirm ein kooperatives Spiel spielen, systematisch, objektiv und reliabel zu erfassen.

An dieser Stelle setzt die Herausforderung von LevelHub an. Um die Methodik weiterzuentwickeln und in neue Kontexte wie Teamentwicklung oder Personalbeurteilung zu übertragen, bedarf es validierter Bewertungsinstrumente.

1.2 Problemstellung

Aktuell mangelt es der Vorgehensweise von LevelHub an objektiven Bewertungsverfahren. Die Einschätzungen beruhen überwiegend auf subjektiven Beobachtungen, was die Vergleichbarkeit einschränkt und die Aussagekraft reduziert.

Vor dem Hintergrund einer geplanten Standardisierung und Skalierung der Methode ist es zentral, objektive Verhaltensankersysteme bereitzustellen – insbesondere bei der unkonventionellen Nutzung von Videospiele im Diagnostik- und Trainingskontext. Nur so lassen sich beobachtete Fähigkeiten zuverlässig und glaubwürdig erfassen.

Verhaltensankersysteme bieten hierfür eine geeignete Grundlage, da sie konkrete Beobachtungskriterien bereitstellen, subjektive Verzerrungen minimieren und die Vergleichbarkeit ermöglichen. Trotz zunehmender Anwendung von GBAs fehlen bislang etablierte Instrumente zur systematischen Erfassung von Verhalten im Kontext kooperativer Videospiele.

1.3 Zielsetzung und Fragestellung

Ziel dieser Arbeit ist die Entwicklung und Überprüfung eines beobachtungsbasierten Verhaltensankersystems (VAS-KOOP) zur Erfassung überfachlicher Fähigkeiten im Kontext eines kooperativen Videospiele. Das Instrument soll es ermöglichen, kooperative Verhaltensweisen systematisch zu beobachten und vergleichbar zu machen, um subjektive Verzerrungen zu reduzieren und die Aussagekraft zu erhöhen. Damit leistet die Arbeit einen Beitrag zur Weiterentwicklung der LevelHub-Methodik und unterstützt die angestrebte Standardisierung.

Als theoretische Grundlage dient das System Non-Technical Skills for Surgeons (NOTSS), das im chirurgischen Setting zur Bewertung nicht-technischer Fähigkeiten eingesetzt wird (Yule, Flin, Paterson-Brown, Maran & Rowley, 2006). Mithilfe der Critical Incident Technique (CIT) nach Flanagan (1954) sollen die Verhaltensindikatoren das bestehende Instrument auf den Kontext eines kooperativen Videospiele angepasst werden.

Die empirische Überprüfung umfasst die Analyse der internen Konsistenz sowie der Interrater-Reliabilität zweier unabhängiger Bewertender. Ergänzend wird die Konstruktvalidität durch den Vergleich von Fremd- und Selbsteinschätzungen geprüft.

Aus der beschriebenen Zielsetzung ergibt sich folgende Forschungsfrage sowie zwei untergeordnete Fragestellungen:

- Lässt sich ein bestehendes Verhaltensankersystem zur Bewertung nicht-technischer Fähigkeiten in adaptierter Form für die Anwendung des VAS-KOOP im Kontext eines kooperativen Videospiele übertragen?
- 1) Wie hoch ist die Interrater-Reliabilität der Fremdeinschätzungen bei der Anwendung des VAS-KOOP im Kontext eines kooperativen Videospiele?
- 2) In welchem Zusammenhang stehen Fremd- und Selbsteinschätzungen bei der Anwendung des VAS-KOOP im Kontext eines kooperativen Videospiele?

Ausgehend von der Zielsetzung und den Fragestellungen wird im nächsten Kapitel der Aufbau der Arbeit skizziert, um die Struktur und den roten Faden der Untersuchung aufzuzeigen.

1.4 Aufbau der Arbeit

Das Kapitel 2 vermittelt den theoretischen Hintergrund und umfasst eine Einführung in die LevelHub-Methodik (2.1) und in Verhaltensankersysteme (2.2). Die Forschungsmethode im Kapitel 3 beginnt mit der Entwicklung des VAS-KOOP (3.1), beschreibt die Durchführung der Datenerhebung (3.2) sowie das methodische Vorgehen bei der Datenauswertung (3.3). Kapitel 4 legt die Ergebnisse dar, darunter deskriptive Daten (4.1), die Prüfung der internen Konsistenz (4.2), die Analyse der Interrater-Reliabilität (4.3) sowie den Vergleich der Fremd- mit den Selbsteinschätzungen (4.4). Im Kapitel 5 folgt die Diskussion mit einer Zusammenfassung (5.1) und Interpretation der Ergebnisse (5.2), wobei die Limitationen der Untersuchung (5.3) diskutiert werden. Kapitel 6 zeigt die Schlussfolgerungen auf, beinhaltet Implikationen für LevelHub (6.1) sowie für praktische Anwendungen (6.2), Perspektiven für zukünftige Forschungen (6.3) und schließt mit einem Fazit (6.4) ab.

2 Theoretische Grundlagen

Im Zentrum dieser Arbeit steht die Frage, wie sich überfachliche Kompetenzen in kooperativen Videospelsituationen strukturiert erfassen lassen – insbesondere mithilfe eines standardisierten Verhaltensankersystems. Um dieses Vorhaben theoretisch zu verorten, führt das Kapitel in zentrale Konzepte ein: Kapitel 2.1 stellt die Initiative LevelHub und deren methodischen Ansatz vor. Kapitel 2.2 beleuchtet die Grundlagen von Verhaltensankersystemen als Instrument zur strukturierten Erfassung überfachlicher Kompetenzen und stellt NOTSS ausführlich vor. Gemeinsam bilden diese theoretischen Grundlagen das Fundament für die Entwicklung und Anwendung des VAS-KOOP.

2.1 LevelHub

LevelHub ist eine sozialunternehmerische Initiative des Netzwerk Grenchen, die sich darauf spezialisiert hat, überfachliche Kompetenzen mithilfe von Videospiele sichtbar und entwickelbar zu machen. Nach erfolgreichen Pilotprojekten im Bereich der Arbeitsintegration wird die Methodik derzeit für weitere Anwendungsfelder wie Personalbeurteilung und Teamentwicklung ausgebaut. Ziel ist es, durch eine Standardisierung die Qualität und Vergleichbarkeit der Beobachtungen zu sichern und so neue Einsatzfelder zu erschliessen.

Kapitel 2.1.1 erläutert den praktischen Ablauf der typischen Vorgehensweise, 2.1.2 erklärt, warum Videospiele für authentische Verhaltensbeobachtungen besonders geeignet sind, und 2.1.3 verortet die Methode theoretisch im Rahmen von GBA.

2.1.1 Vorgehensweise

Die LevelHub-Methodik nutzt konventionelle Videospiele, um Verhaltensmuster auf unkonventionelle Weise sichtbar zu machen. Die Abbildung 1 zeigt in modellhaft, wie sich ein beispielhafter Tag gestaltet.

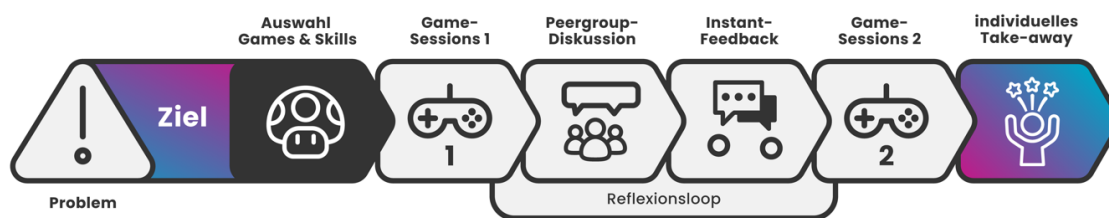


Abbildung 1. Ablauf eines typischen LevelHub-Tages: Vom definierten Ziel über die Auswahl passender Spiele, Spielphasen und Reflexionen bis hin zum individuellen Take-away (LevelHub).

Teilnehmende spielen ein kooperatives Videospiele, das je nach Ziel unterschiedliche Fähigkeiten erfordert. Beispielsweise kann ein Spiel in einer Gruppe kollaborative Problemlöseprozesse, Entscheidungsfindung sowie Führungsverhalten sichtbar machen. Eine erste Game-Session dient dem Kennenlernen der Steuerung und der Mitspielenden. Mitarbeitende von LevelHub nehmen dabei eine beobachtende Rolle ein und notieren Auffälligkeiten (siehe Abbildung 2).

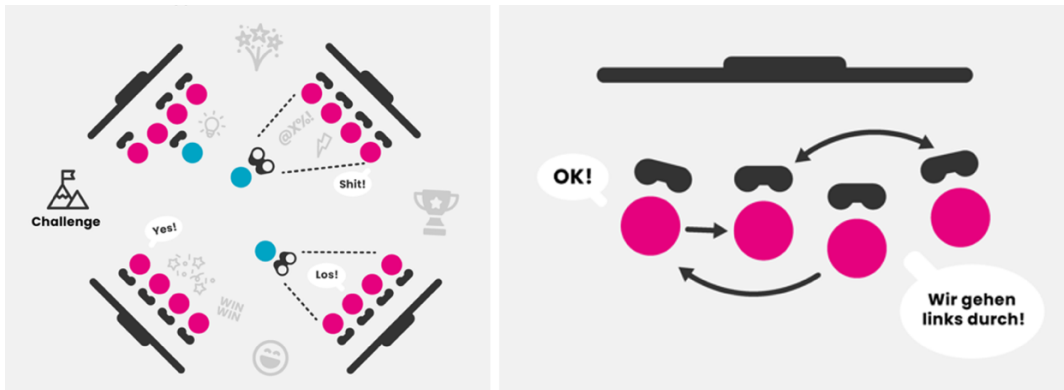


Abbildung 2. Links: Modellhafte Darstellung einer klassischen Game-Session, wobei mehrere Teams gleichzeitig spielen und Beobachtungen gemacht werden. Rechts: eine symbolische Darstellung eines erfolgreichen kooperativen Prozesses (LevelHub).

Anschauliche Situationen werden danach im Plenum aufgenommen und bilden in Peergroup-Diskussionen oder Instant-Feedback die gemeinsame Grundlage für den Reflexionsloop. Es werden beobachtete Verhaltensmuster angesprochen und Strategien erarbeitet, wie damit umgegangen werden könnte. In der zweiten Game-Session sollen persönliche oder teambasierte Strategien direkt ausprobiert und geprüft werden. In vielen Fällen können durch die spielerisch ungezwungene Situation Erfolgserlebnisse erlebt werden, was für weitere Selbstreflexionsprozess förderlich sein kann. Kernelement der Methodik ist der anschließende Transfer in den Berufsalltag, um die persönlichen Erkenntnisse nachhaltig zu festigen.

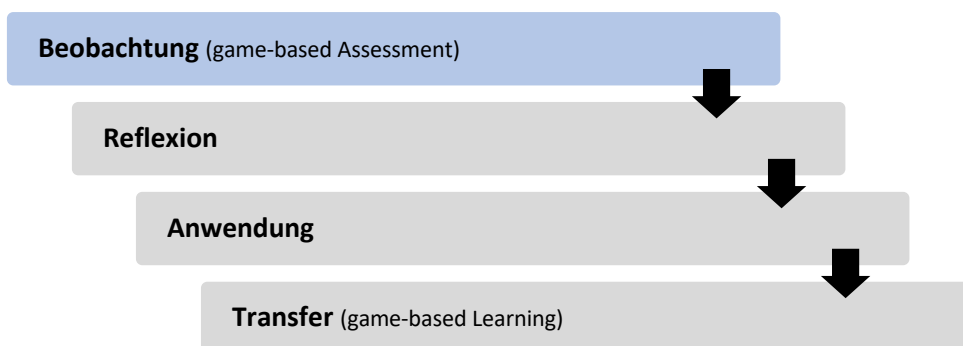


Abbildung 3. Vierstufiges Wirkmodell der Vorgehensweise – von der Beobachtung über Reflexion und Anwendung bis zum Transfer in den Alltag (LevelHub).

Die Beobachtung stellt dabei den Ausgangspunkt eines mehrstufigen Lernprozesses (Abbildung 3) dar, in dem beispielsweise kooperative Verhaltensweisen zunächst sichtbar gemacht und anschliessend durch Reflexion, Anwendung und Transfer weiterentwickelt werden. Während in der vorliegenden Untersuchung der Fokus auf der Beobachtungsebene liegt, wird im Kapitel 2.1.3 detaillierter auf GBAs eingegangen.

2.1.2 Warum Videospiele?

Obwohl Gaming häufig mit negativen Stereotypen assoziiert wird, zeigen Studien zunehmend positive Effekte, insbesondere im Hinblick auf überfachliche Fähigkeiten (Bailey, Pearson, Gkatzidou & Green, 2006; Beigi et al., 2025; Connolly et al., 2012; Granic et al., 2014; Trespalacios, Chamberlin & Gallagher, 2011). Professionell entwickelte Videospiele bieten durch ihre anspruchsvoll gestalteten virtuellen Welten und interaktive Steuerung die Möglichkeit tiefer Immersion. In diesem Zustand tritt das Bewusstsein in den Hintergrund, während die Spielenden ganz im Geschehen aufgehen (Sweetser & Wyeth, 2005).

Durch die kognitive Beanspruchung und den entstehenden Flow-Zustand (Csikszentmihalyi, 1990) stehen nur begrenzt mentale Ressourcen zur Verfügung, um die eigene Wirkung auf andere gezielt zu steuern (Goffman, 1959). Soziale Erwünschtheit tritt zurück, was nach Einschätzung durch LevelHub authentische Verhaltensbeobachtungen ermöglicht, die für spätere Reflexionen besonders aufschlussreich sein können. Zudem steigern interaktive Lernumgebungen nachweislich das Engagement (Moreno & Mayer, 2007) und damit die Bereitschaft zur Selbstreflexion – ein zentrales Element der LevelHub-Methodik.

Videospiele rücken zunehmend als diagnostisches Instrument in den Fokus. Im folgenden Kapitel wird der Einsatz von GBA näher betrachtet.

2.1.3 Game-based Assessment

GBA bezeichnet die systematische Erfassung von Kompetenzen, Fähigkeiten oder Persönlichkeitsmerkmalen anhand von beobachtbarem Verhalten in interaktiven Spielsituationen. Im Unterschied zu klassischen Tests oder Fragebögen erfolgt die Datenerhebung nicht durch Selbstberichte, sondern durch die Analyse von Handlungen innerhalb eines Spiels (Landers, 2015). Wichtig ist die Abgrenzung von verwandten Konzepten: Während game-based Learning (GBL) den Fokus auf das Entwickeln von Kompetenzen durch Spiele legt, also didaktische Inhalte vermittelt, zielt GBA auf die diagnostische Auswertung von Verhalten. Auch Gamification unterscheidet sich von GBA, da sie lediglich einzelne Spielelemente in spielfremde Kontexte einbettet, ohne ein vollständiges Spiel zu nutzen (Landers & Sanchez, 2022).

In der Praxis wird GBA insbesondere im Personalbereich eingesetzt (Ellison, Johnson, Tomczak, Siemsen & Gonzalez, 2020; Ohlms, Melchers & Kanning, 2024). Auch in schulischen und universitären Kontexten finden sich Anwendungen, beispielsweise zur Bewertung von Soft Skills (Gomez et al., 2023).

GBA gilt als besonders motivierend, da das Spielen in der Regel als weniger bedrohlich empfunden wird als klassische Testsituationen. Studien zeigen, dass sich dies positiv auf die Akzeptanz der Verfahren auswirkt (Landers, 2015). Darüber hinaus erlaubt die Spielsituation eine reichhaltige Datenerfassung. Neben den Endresultaten können Interaktionshäufigkeiten oder Kommunikationsverhalten ausgewertet werden (Bhatia & Ryan, 2018).

Allerdings ist die Umsetzung komplex. Eine valide und faire Konstruktion erfordert sorgfältige Testentwicklung, fundierte psychometrische Analysen und den Nachweis

kontextbezogener Validität (Ellison et al., 2020). Zudem ist die Entwicklung eigener Spiele mit hohen technischen und finanziellen Anforderungen verbunden, sodass die Nutzung konventioneller Spiele einen Vorteil sein kann (Landers, 2015).

Für die vorliegende Arbeit ist GBA zentral, weil kooperative Videospiele als Beobachtungsszenario genutzt werden, um überfachliche Fähigkeiten wie Teamarbeit oder Entscheidungsfindung zu erfassen. Dabei wird das beobachtbare Verhalten durch ein Verhaltensankersystem strukturiert ausgewertet.

2.2 Verhaltensankersysteme

Die Bewertung individueller Fähigkeiten innerhalb der LevelHub-Methodik – insbesondere in simulationsbasierten Gruppenprozessen – erfordert strukturierte und objektive Verfahren. Gerade soziale und kognitive Kompetenzen wie Teamfähigkeit, Kommunikationsstärke oder Problemlösungsvermögen lassen sich nicht direkt messen.

Verhaltensankersysteme haben sich hier als wirksames Instrument etabliert. Sie ermöglichen es, latente Merkmale über beobachtbares Verhalten systematisch zu erschliessen (Dietz et al., 2014; Manser, Schmutz & Perry, 2018). Dabei wird das Verhalten nicht willkürlich, sondern unter kontrollierten Bedingungen beobachtet, um fundierte Aussagen über relevante Fähigkeiten treffen zu können (Höft & Kersting, 2018).

2.2.1 Definition

Ein Verhaltensankersystem ist ein strukturiertes Beobachtungsinstrument, das spezifische, beobachtbare Verhaltensindikatoren nutzt, um soziale und kognitive Kompetenzen objektiv und nachvollziehbar zu bewerten (Ellison et al., 2020; Ohlms et al., 2024). Diese Indikatoren fungieren als Ankerpunkte auf einer Skala, die typischerweise Verhalten mit niedriger, mittlerer und hoher Ausprägung beschreibt (Kleinmann, 2016).

Der Vorteil: Die kognitiven Anforderungen an Beobachtende werden reduziert, und der Beobachtungsprozess wird strukturierter (Hennessy, Mabey & Warr, 1998). Besonders in Gruppensettings eignen sich diese Systeme gut, um Teamfähigkeit objektiver zu erfassen – vorausgesetzt, die Beobachtenden sind geschult (Flin & Martin, 2001; Obermann, 2002). Die Objektivität kann zusätzlich erhöht werden, wenn beispielsweise Psycholog:innen in die Beobachtung eingebunden werden (Kleinmann, 2016).

Dank ihrer Spezifität ermöglichen Verhaltensankersysteme ein standardisiertes Vokabular für Bewertung und Rückmeldungen. Damit wird auch eine differenzierte Nachbesprechung möglich – mit klaren Beschreibungen der Dimensionen (Fletcher et al., 2003; Yule, Flin, Paterson-Brown & Maran, 2006).

2.2.2 Evidenzbasis und Anwendungsbereiche

Verhaltensankersysteme verfügen insbesondere im Gesundheitsbereich über eine gut dokumentierte empirische Fundierung. Laut einer umfassenden Übersichtsarbeit von Dietz et al. (2014) weisen 75 % der analysierten Systeme Nachweise zur Reliabilität und 70 % zur Validität

auf. Dabei wird betont, dass sich die Gültigkeit eines Instruments nicht generalisieren lässt, sondern kontextbezogen zu bewerten ist – abhängig vom jeweiligen Anwendungsfeld und der Zielsetzung.

In der klinischen Praxis konnten verschiedene Studien die Eignung dieser Systeme zur strukturierten Erfassung nicht-technischer Kompetenzen bestätigen. So belegen Untersuchungen zur Konstruktvalidität, Anwendbarkeit und Beobachtungsgenauigkeit, dass Verhaltensankersysteme auch in komplexen, teamintensiven Situationen verlässliche Ergebnisse liefern – sowohl in simulierten als auch in realitätsnahen Settings (Fletcher et al., 2003; Mitchell et al., 2012; Yule et al., 2018).

Neben dem medizinischen Bereich finden Verhaltensankersysteme auch in anderen sicherheitskritischen Bereichen Anwendung – etwa in der Luftfahrt (Flin & Martin, 2001) oder der Schifffahrt (da Conceição, Mendes, Teodoro & Dahlman, 2019). In jüngerer Zeit nimmt auch der Einsatz in kooperativen Videospiele zu, wobei besonders Teamarbeit simuliert werden soll (Farah, Dorneich & Gilbert, 2022). Ebenso sind Verhaltensankersysteme fester Bestandteil eignungsdiagnostischer Verfahren wie Assessment- und Development-Center, wo sie zur Beurteilung überfachlicher Fähigkeiten herangezogen werden (Kleinmann, 2016; Obermann, 2002).

Aufgrund ihrer breiten Anwendbarkeit und empirischen Fundierung eignen sich Verhaltensankersysteme auch für innovative Einsatzbereiche wie im Kontext kooperativer Videospiele. Im folgenden Kapitel wird deshalb das NOTSS-System (Yule et al., 2018) als Referenzmodell näher betrachtet, das für die vorliegende Studie adaptiert wurde.

2.2.3 Non-Technical Skills for Surgeons

Für die vorliegende Arbeit wurde ein Verhaltensankersystem basierend auf dem validierten Beobachtungssystem NOTSS (Yule et al., 2018) entwickelt. Das ursprüngliche System wurde in Schottland von einem interdisziplinären Team aus chirurgischem, psychologischem und anästhetischem Fachpersonal entwickelt, um die beobachtbaren Verhaltensaspekte im Operationssaal systematisch zu erfassen. Es wurde anfänglich als strukturiertes Feedbackinstrument für chirurgische Trainingssituationen genutzt und ermöglicht eine gezielte Rückmeldung zu beobachtbarem Verhalten in realen oder simulierten Situationen (Yule, Flin, Paterson-Brown & Maran, 2006).

Nach der Prüfung ähnlicher Systeme wie Observational Teamwork Assessment for Surgery (Undre, Sevdalis, Healey, Darzi & Vincent, 2007) oder Scrub Practitioners' List of Intraoperative Non-Technical Skills System (Mitchell et al., 2012) wurde NOTSS (Yule, Flin, Paterson-Brown, Maran, et al., 2006) aufgrund seiner Übersichtlichkeit, seiner passenden Dimensionen zur Spielsituation sowie der klar definierten und beobachtbaren Verhaltensmarker ausgewählt. Während viele der geprüften Systeme auf Effektivität von Teams eingestellt sind oder zu viele Dimensionen gleichzeitig messen, bietet NOTSS eine generischere und dennoch valide Grundlage (Yule et al., 2018) für die Übertragung auf die simulierte kooperative Spielsituationen. Die Auswahl des Systems wurde getroffen, weil sie die kooperativen Anforderungen in

Spielsettings gut abbilden und gleichzeitig zentrale Aspekte überfachlicher Fähigkeiten beinhalten.

Die Taxonomie (Abbildung 4) gliedert nicht-technische Fähigkeiten in vier Hauptkategorien: Situationsbewusstsein, Entscheidungsfindung, Kommunikation & Teamarbeit sowie Führung. Jede Kategorie umfasst spezifische Dimensionen. Situationsbewusstsein beinhaltet das Sammeln und Verstehen von Informationen sowie die antizipierende Einschätzung zukünftiger Entwicklungen. Die Entscheidungsfindung umfasst das Abwägen von Handlungsoptionen, die Auswahl und Kommunikation einer Option sowie deren Umsetzung und Reflexion. Kommunikation & Teamarbeit beziehen sich auf den Informationsaustausch, das Schaffen eines gemeinsamen Verständnisses sowie die Koordination von Teamaktivitäten. Die Kategorie Führung beinhaltet das Setzen und Einhalten von Standards, die Unterstützung anderer sowie den konstruktiven Umgang mit Drucksituationen.

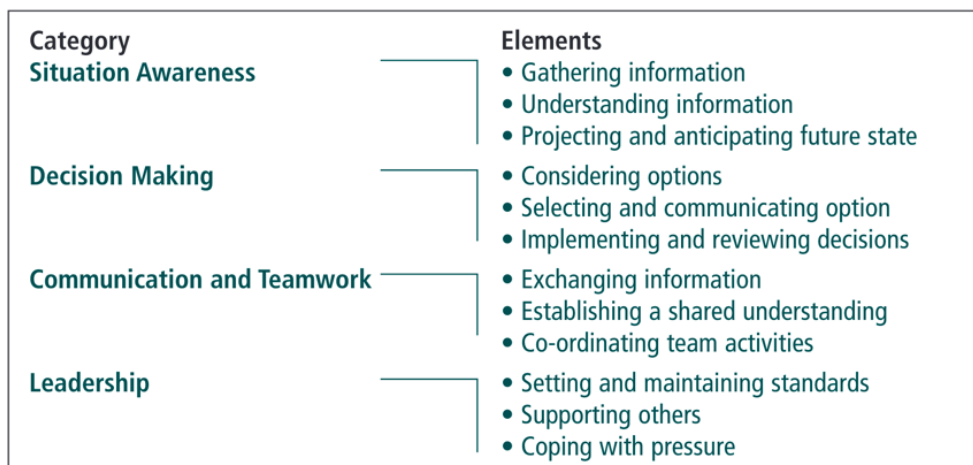


Abbildung 4. Originale Skill-Taxonomie des NOTSS (Yule, Flin, Paterson-Brown, Maran & Rowley, 2006).

Für die Bewertung der nicht-technischen Fähigkeiten kam eine vierstufige Skala zum Einsatz: *schwach* (1), *grenzwertig* (2), *akzeptabel* (3), *gut* (4). Sie diente der strukturierten Einschätzung beobachtbarer Verhaltensweisen auf Dimensions- und Kategorieebene. Die Skala ermöglicht eine differenzierte und vergleichbare Bewertung von nicht-technischen Fähigkeiten (Yule, Flin, Paterson-Brown & Maran, 2006).

Die Anwendung des NOTSS wurde ursprünglich im OP-Kontext zur Fremdbeobachtung genutzt, eignet sich aber laut Yule et al. (2018) auch für Forschung und Bildung, um nicht-technische Fähigkeiten systematisch zu erfassen. In der Anwendung zeigt sich jedoch, dass eine Einführung in die Systematik sowie praktische Übung erforderlich sind, um valide und vergleichbare Einschätzungen treffen zu können. Dies gilt insbesondere für die zuverlässige Beurteilungen durch geschulte Bewertende. Die klare Struktur des Systems unterstützt sowohl Diskussionen als auch Selbsteinschätzung (Yule, Flin, Paterson-Brown & Maran, 2006). Studien zeigen zudem, dass chirurgische Fachpersonen zwar in der Lage sind, ihre technischen Kompetenzen einzuschätzen, ihre nicht-technischen Fähigkeiten jedoch deutlich schwerer

selbstbewerten können (Arora et al., 2011). Dies spricht für eine kontinuierliche Verbesserung solcher strukturierten Verhaltensankersysteme.

Aufgrund der breiteren Anwendbarkeit im Kontext psychologischer Forschung und Bildung wird im weiteren Verlauf dieser Arbeit der Begriff überfachliche Fähigkeiten anstelle von nicht-technischen Fähigkeiten verwendet. Für die Entwicklung des Verhaltensankersystem im Kontext eines kooperativen Videospiele wurden leichte Anpassungen an der Taxonomie vorgenommen. Im Kapitel 3.1.2 wird die Adaption des Modells ausführlicher dargestellt.

3 Methode

Zur Beantwortung der Forschungsfrage wurde ein mehrstufiges methodisches Vorgehen gewählt, das sowohl die Entwicklung als auch die Überprüfung eines neuen Verhaltensankersystems umfasst. Ziel war die Entwicklung eines Instrumentes, das überfachliche Fähigkeiten im Kontext eines kooperativen Videospiele systematisch und nachvollziehbar erfassen kann.

Das vorliegende Kapitel gliedert sich in drei Teile: Zunächst wird die Entwicklung des VAS-KOOP beschrieben (Kapitel 3.1), gefolgt von der Dokumentation der Datenerhebung (Kapitel 3.2) und der geplanten statistischen Auswertung (Kapitel 3.3).

3.1 Entwicklung des VAS-KOOP

Im Rahmen dieses Schritts wurde ein bestehendes, validiertes Beobachtungssystem aus dem medizinischen Kontext namens NOTSS (Yule, Flin, Paterson-Brown, Maran, et al., 2006) analysiert und gezielt für den Einsatz in kooperativen Spielszenarien weiterentwickelt. Kapitel 3.1.1 beschreibt die Auswahl geeigneter Modelle und Fähigkeiten. In Kapitel 3.1.2 werden die systematische Modifikation des Instruments sowie die Entwicklung kontextangepasster Verhaltensindikatoren dokumentiert. Im abschliessenden Kapitel 3.1.3 wird das finale Instrument VAS-KOOP mit seinen Anwendungsformen vorgestellt.

3.1.1 Recherche zu passenden Modellen und Fähigkeiten

Zur Entwicklung eines geeigneten Verhaltensankersystems für kooperative Spielsituationen wurde zunächst eine umfassende Recherche zu bestehenden Verhaltensankersystemen durchgeführt. Ziel war es, ein etabliertes Modell zu identifizieren, das sich sowohl methodisch bewährt hat als auch jene Fähigkeiten abbildet, die in der geplanten Spielsituation relevant sind. Grundlage hierfür bildete ein exploratives Gespräch mit Prof. Dr. Tanja Manser, Direktorin Hochschule für Angewandte Psychologie FHNW, die als Expertin im Bereich der Verhaltensbeobachtung in Gesundheitsberufen gilt. Das Gespräch diente der ersten Orientierung im Themenfeld sowie der Vermeidung häufiger methodischer Fehler bei der Systementwicklung.

Im Zentrum der Recherche stand die Frage, welche Fähigkeiten in kooperativen Spielsituationen besonders relevant sind. Dabei zeigten sich insbesondere folgende Aspekte im Spielverhalten von Bedeutung: Informationsverarbeitung (z. B. das Teilen und Bestätigen von Informationen), Aufgabenmanagement (z. B. Priorisierung und Koordination), Stress- und Konfliktmanagement, Führung, Unterstützungsverhalten sowie der Umgang mit Fehlern. Weitere übergeordnete Fähigkeiten wie Anpassungsfähigkeit und Durchsetzungsvermögen wurden zunächst ebenfalls als wichtig betrachtet. Gleichzeitig musste darauf geachtet werden, die Anzahl beobachteter Dimensionen zu begrenzen, um eine praktikabel und valide Bewertung zu gewährleisten.

Im Rahmen der Literaturanalyse wurden verschiedene etablierte Verhaltensankersysteme geprüft (da Conceição et al., 2019; Mitchell et al., 2012; Undre et al., 2007). Nach eingehender Analyse sowie Rücksprache mit dem Praxispartner fiel die Wahl auf NOTSS (Yule, Flin,

Paterson-Brown, Maran, et al., 2006), das ursprünglich für den chirurgischen Kontext entwickelt wurde. Es zeichnete sich jedoch durch eine klare, übersichtliche Struktur und potenzielle Übertragbarkeit auf das spielerische Setting aus. Andere geprüften Systeme fokussierten entweder stärker auf Effektivität von Teams oder erfassten zu viele Dimensionen gleichzeitig.

Die Auswahl des Systems wurde getroffen, da es die kooperativen Anforderungen in Spielsettings gut abbildet und zentrale Aspekte zukunftsrelevanter Fähigkeiten (z.B. Kommunikation und Kollaboration) beinhaltet.

3.1.2 Modifikation eines etablierten Instruments

Im Rahmen dieser Arbeit wurde NOTSS (Yule, Flin, Paterson-Brown, Maran, et al., 2006) übersetzt (Anhang A) und gezielt auf die Anforderungen eines kooperativen Videospieles angepasst. Das Ziel war, ein Instrument zur Erfassung überfachlicher Fähigkeiten im Spielkontext zu entwickeln.

Anpassung der Kategorien und Dimensionen

Das ursprüngliche NOTSS (Yule, Flin, Paterson-Brown, Maran, et al., 2006) umfasste vier Hauptkategorien: *Situationsbewusstsein*, *Entscheidungsfindung*, *Kommunikation & Teamarbeit* sowie *Führung*. Die Kategorie *Situationsbewusstsein* wurde für den Einsatz in diesem Projekt bewusst ausgeschlossen. Zum einen erschien das Beobachtungssystem mit vier Kategorien zu umfangreich für eine valide Anwendung im Spielsetting. Zum anderen bestand die Annahme, dass insbesondere spielerfahrene Personen in dieser Kategorie signifikant höhere Bewertungen erhalten würden, was zu Verzerrungen im Vergleich zu weniger spielerfahrenen Teilnehmenden führen könnte. Der Fokus wurde daher auf die drei verbleibenden Kategorien gelegt.

Innerhalb der Kategorie *Führung* wurde zudem die ursprüngliche Dimension *Festlegen und Einhalten von Standards* durch die Dimension *Ermütigung anderer* ersetzt. Diese Entscheidung basiert auf Überlegungen aus der Theorie der transformationalen Führung, nach der das Fördern und Ermütigen von Teammitgliedern – auch in spielerischen Kontexten – als wirksames Führungsverhalten gilt (Podsakoff, MacKenzie, Moorman & Fetter, 1990; Schaubroeck, Lam & Cha, 2007). Die übrigen Dimensionen blieben inhaltlich unverändert, sodass sich das angepasste Verhaltensankersystem nun wie in Tabelle 1 gliedert.

Tabelle 1

Adaptierte Taxonomie mit den drei Hauptkategorien und jeweils drei Unterdimensionen

Entscheidungsfindung	Kommunikation & Teamarbeit	Führung
Betrachten von Optionen	Informationsaustausch	Ermütigung anderer
Auswahl und Kommunikation einer Option	Gemeinsames Verständnis herstellen	Unterstützung anderer
Umsetzung und Überprüfung von Entscheidungen	Koordinierung von Teamaktivitäten	Umgang mit Druck

Entwicklung der Verhaltensindikatoren mittels CIT

Ein zentraler Bestandteil der Modifikation war die Entwicklung kontextspezifischer Verhaltensindikatoren für jede der neun Dimensionen. Hierzu wurde die CIT nach Flanagan (1954) angewendet – ein bewährtes, offenes Verfahren zur Anforderungsanalyse, das insbesondere in der Personalpsychologie eingesetzt wird. Ziel der CIT ist es, erfolgskritische Verhaltensweisen zu identifizieren, die zum Gelingen oder Misslingen einer bestimmten Tätigkeit beitragen. Die Technik erlaubt sowohl retrospektive als auch prospektive Betrachtungen und eignet sich zur Ableitung beobachtbarer Indikatoren für bestimmte Fähigkeiten.

Der Ablauf folgte den fünf von Flanagan (1954) vorgeschlagenen Schritten:

- Zielbestimmung (1): Im Spielkontext galt das Erreichen gemeinsamer Spielziele (z. B. Sterne sammeln) sowie die Förderung von Spielfreude.
- Planung (2): Es wurde ein Beobachtungsworkshop vorbereitet, Fragen entwickelt, Personen eingeladen und Rollen verteilt.
- Datenerhebung (3): Beobachtende sammelten positive und negative Verhaltensbeispiele während der Spielsituationen im Spiel *Overcooked! 2* (Team17 & Ghost Town Games, 2018) mithilfe von Reflexionsfragen auf einer PowerPoint-Vorlage.
- Datenanalyse (4): Die gesammelten Verhaltensbeispiele wurden systematisch den angepassten Dimensionen zugeordnet.
- Deutung und Bericht (5): Die Ergebnisse flossen in die Interpretation im Diskussionsteil (Kapitel 5.3) ein.

Der CIT-Workshop wurde am 18. Februar 2025 im LevelHub durchgeführt. Trotz kurzfristiger Absagen konnte der Workshop mit drei Spielenden und zwei Beobachtenden umgesetzt werden. Die Spielphasen wurden bewusst strukturiert: Nach je einem kurzen Warm-up-Level wurden nacheinander die Dimensionen Entscheidungsfindung, Kommunikation & Teamarbeit sowie Führung fokussiert beleuchtet. Zwischen den Levels erfolgten jeweils 4-minütige Gruppendiskussionen, in denen Strategien zur Leistungssteigerung erarbeitet wurde. Die beobachteten Verhaltensweisen wurden mithilfe von Leitfragen (Anhang B) gemeinsam, auf einer Matrix als *positiv* oder *negativ* zugeordnet und als Grundlage für die Adaption der Verhaltensindikatoren dokumentiert (siehe Abbildung 5).

	-	1	2	3	4	5	+
Ich ignorierte Probleme, hörte nicht auf Vorschläge und schlug keine Lösungen vor.							Ich erkannte und erwähnte Probleme, schlug alternative Wege vor, diskutierte Optionen mit dem Team und fragte nach Meinungen.
Ich gab Infos nicht weiter, kommunizierte keinen Plan, traf unklare oder widersprüchliche Entscheidungen und schloss andere Optionen aus.							Ich gab klare Anweisungen, kommunizierte und erklärte Entscheidungen, wiederholte den Plan verständlich und fragte nach Bestätigung.
Ich setzte den Plan nicht um, sprach Probleme nicht an, blieb stur bei schlechten Entscheidungen, wiederholte Fehler, reflektierte diese nicht kritisch und war überstürzt.							Ich setzte den Plan um, sprach Probleme direkt an, korrigierte den Prozess bei Veränderungen, informierte das Team über den Fortschritt und rief Unterstützung bei Bedarf.

Abbildung 6. Ausschnitt aus dem VAS-KOOP zur Selbsteinschätzung mit angepasster Bewertungsskala (eigene Abbildung).

Durch die gezielte Reduktion und Adaption der NOTSS-Kategorien sowie die Entwicklung kontextspezifischer Verhaltensindikatoren mittels CIT entstand ein praxisnahes Beobachtungsinstrument für kooperative Spielsituationen. Im folgenden Kapitel wird das finale Verhaltensankersystem VAS-KOOP in seiner Gesamtheit vorgestellt.

3.1.3 Verhaltensankersystem VAS-KOOP

Das entwickelte Verhaltensankersystem VAS-KOOP ist für die Anwendung in Forschung, Workshops und eignungsdiagnostischen Kontexten gedacht – speziell im Rahmen kooperativer Videospiele wie *Overcooked! 2* (Team17 & Ghost Town Games, 2018). Es dient der systematischen Beobachtung, dem Vergleich von Verhaltensweisen sowie der anschließenden Reflexion überfachlicher Fähigkeiten.

Das System liegt in zwei Versionen vor: einer Fremdeinschätzung durch geschulte Beobachtende sowie einer Selbsteinschätzung durch die Proband:innen. Diese Dualität erlaubt eine methodische Triangulation, durch die ergänzende Perspektiven auf das beobachtete Verhalten gewonnen werden können. Ziel ist es, neben der externen Beobachtungen auch das subjektive Erleben der Proband:innen zu erfassen und beide Datenquellen miteinander in Beziehung zu setzen.

Entscheidungsfindung							
Fähigkeiten zur Diagnose der Situation und zur Entscheidungsfindung, um einen geeigneten Handlungsverlauf zu bestimmen.							
Betrachten von Optionen							
Ignoriert Probleme; hört nicht auf Vorschläge; schlägt keine Lösungen vor							Erkennt und erwähnt Probleme; schlägt alternative Wege vor; diskutiert Optionen mit dem Team; fragt nach Meinungen
Auswahl und Kommunikation einer Option							
Gibt Infos nicht weiter; kommuniziert keinen Plan; unklare oder widersprüchliche Entscheidungen; schliesst andere Optionen aus							Gibt klare Anweisungen; kommuniziert und erklärt Entscheidungen; wiederholt den Plan verständlich; fragt nach Bestätigung
Umsetzung und Überprüfung von Entscheidungen							
Setzt Plan nicht um; spricht Probleme nicht an; bleibt stur bei schlechten Entscheidung; wiederholt Fehler; reflektiert diese nicht kritisch, wirkt überstürzt							Setzt den Plan um; spricht Probleme direkt an; korrigiert Prozess bei Veränderungen; informiert Team über Fortschritt; ruft Unterstützung bei Bedarf

Abbildung 7. Ausschnitt aus dem VAS-KOOP zur Fremdeinschätzung mit Kategorien, Definitionen sowie positiven und negativen Verhaltensindikatoren pro Dimensionen (eigene Abbildung).

Das VAS-KOOP umfasst neun Dimensionen, aufgeteilt auf drei Kategorien überfachlicher Fähigkeiten: *Entscheidungsfindung*, *Kommunikation & Teamarbeit* sowie *Führung* (siehe Tabelle 1). Für jede Dimension wurden sowohl *negative* (1) als auch *positive* (2) Verhaltensindikatoren definiert (siehe Abbildung 4). Diese Anker wurden auf Basis der CIT abgeleitet und in strukturierter Form in eine 5-stufige Skala überführt. Die Skalenpole sind explizit formuliert, um Bewertung des beobachteten Verhaltens möglichst objektiv und nachvollziehbar zu machen.

Die Selbsteinschätzung (siehe Abbildung 3) erfolgt direkt im Anschluss an die kooperative Spielsituation. Dabei wurden die Verhaltensanker in Ich-Formulierungen übersetzt, um eine intuitive und klare Selbsteinschätzung zu ermöglichen. Die Proband:innen füllen den Bogen individuell aus, ohne Rücksprache mit anderen.

Die Fremdeinschätzung (siehe Abbildung 4) wird hingegen retrospektiv anhand von Videomaterial vorgenommen. Die Beobachtenden tragen ihre Bewertungen direkt in eine vorbereitete Excel-Matrix ein, in der alle Dimensionen, Kategorien und Sitzplätze zugeordnet sind. Die Fremdeinschätzung beinhaltet zusätzlich Definitionshilfen und strukturierte Kategorien, um eine möglichst konsistente Bewertung sicherzustellen.

Das vollständige VAS-KOOP – sowohl für Selbst- als auch Fremdeinschätzung – ist im Anhang C (Selbsteinschätzung) und Anhang D (Fremdeinschätzung) dokumentiert.

3.2 Datenerhebung

Zur Erprobung des entwickelten Verhaltensankersystem VAS-KOOP wurde ein methodisch kombinierter Zugang gewählt. Die Datenerhebung fand im Rahmen des Videospiele *Overcooked! 2* (Team17 & Ghost Town Games, 2018) auf einer Nintendo Switch statt. Dabei spielten Gruppen à drei bis vier Personen gemeinsam, diskutierten und füllten anschließend eine Selbsteinschätzung aus. Ziel war eine beobachtungs-basierte Erfassung überfachlicher Fähigkeiten.

Das Kapitel gliedert sich in drei Abschnitte: Zunächst wird die Stichprobe (3.2.1) beschrieben, gefolgt vom konkreten Ablauf der Datenerhebung (3.2.2), in welchem die Durchführung der Sessions detailliert dargestellt wird. Der dritte Abschnitt (3.2.3) widmet sich der videogestützten Beobachtungsanalyse, die im Anschluss an die Erhebung zur Bewertung eingesetzt wurde.

3.2.1 Stichprobe

Die angestrebte Stichprobe umfasst 30 volljährige Personen mit grundlegenden Deutschkenntnissen. Die Rekrutierung erfolgt über das Netzwerk Grenchen, das diverse Angebote für Arbeitsintegrations-Massnahmen anbietet. Vier Abteilungsleitende wurden mithilfe von Informationsmaterial (Anhang E) gebeten, Gruppen mit drei bis vier Personen für die Teilnahme an festgelegten Zeitfenstern zu organisieren. Dafür wurde ein Terminplan (Anhang F) mit freien Zeiträumen vorbereitet, in dem die Zuweisenden einschreiben konnten, wann und wie viele Proband:innen teilnehmen würden.

Der grösste Teil der Stichprobe bestand aus Teilnehmenden von Arbeitsintegrations-Programmen. Zusätzlich nahmen auch festangestellte Mitarbeitende des Netzwerk Grenchen an der Erhebung teil. Gruppen mit weniger als drei Personen wurden von der Teilnahme ausgeschlossen, da das untersuchte Verhalten in einem interaktiven Gruppensetting stattfinden musste. Es wurden keine weiteren Einschluss- oder Ausschlusskriterien definiert, um eine möglichst vielfältige Stichprobe zu erhalten. Zusätzliche personenbezogenen Daten wie Alter oder Geschlecht wurden nicht erhoben, da diese für das Forschungsinteresse der Studie nicht relevant waren.

3.2.2 Ablauf der Datenerhebung

Die Datenerhebung erfolgte in mehreren klar strukturierten Schritten (siehe Abbildung 8). Der Ablauf in vollständiger Form ist im Anhang G ersichtlich. Ziel war es, mit dem entwickelten Beobachtungsinstrument VAS-KOOP kooperative Verhaltensweisen in einer gemeinsamen Spielsituation zu erfassen.

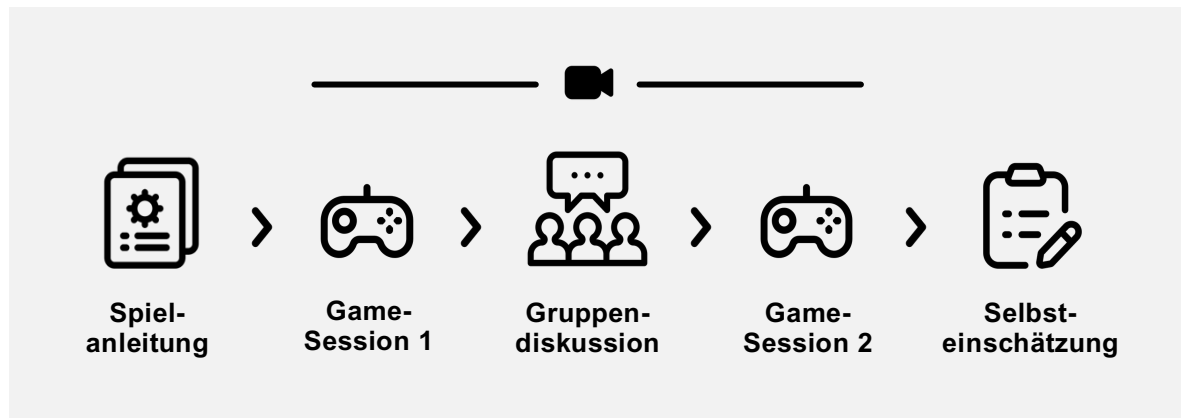


Abbildung 8. Ablauf der Datenerhebung mit den zentralen und videodokumentierten Phasen (eigene Abbildung).

Dabei wurde betont, dass es sich um die Erprobung eines neuen Instruments handelt und alle Daten vollständig anonymisiert erhoben werden – eine Rückführung auf einzelne Personen ist somit ausgeschlossen. Im Zentrum stand das beobachtbare Verhalten innerhalb der Gruppe. Um eine entspannte Atmosphäre zu schaffen, wurde ausdrücklich darauf hingewiesen, dass es sich nicht um einen Leistungstest handelt und sich die Proband:innen möglichst natürlich verhalten sollen.

Zu Beginn wurde den Proband:innen das Ziel der Studie erklärt. Danach erfolgte die Übergabe der Einverständniserklärung, die von allen nach dem Lesen unterzeichnet wurde. Anschliessend wurde der Ablauf der gesamten Erhebung erläutert: Die Proband:innen wurden informiert, dass zwei Game-Sessions mit einer dazwischenliegenden kurzen Gruppendiskussion stattfinden würden. Zusätzlich wurde die Selbsteinschätzung am Ende erwähnt und auf die Videoaufnahme während der Spielphasen hingewiesen. Offene Fragen konnten an dieser Stelle geklärt werden.

Vor Beginn des Spiels wurde die Aufgabe anhand Abbildung 9 erklärt: Gemeinsam sollte in einer kooperativen Spielsituation eine Suppe mit drei Zwiebeln gekocht werden. Dazu gehörten das Holen und Schneiden von Zwiebeln, das Kochen der Suppe und das anschließende Servieren. Vor Spielbeginn wurde die Steuerung des Spiels erklärt, insbesondere die Funktion der verschiedenen Knöpfe auf dem Controller.



Abbildung 9. Leveldesign des durchgeführten Level 1 in Overcooked! 2 (Team17 & Ghost Town Games, 2018) mit Darstellung sämtlicher Utensilien, Bestellungen, Punktestand und Zeitangaben (Screenshot).

Anschließend wurde die Videoaufnahme gestartet. Während der 1. Game-Session lernten die Proband:innen das Spiel und dessen Funktionen kennen, wobei sie sich bei Fragen melden konnten. Danach folgte eine etwa vierminütige Gruppendiskussion. In dieser Zeit sollten die Proband:innen gemeinsam Strategien entwickeln, um die Spielaufgabe gemeinsam effizienter zu bewältigen – mit dem Ziel, in der nächsten Runde mehr Punkte zu erreichen. Im Anschluss startete die 2. Game-Session, in der die Gruppe erneut das gleiche Level absolvierte. Nach Beendigung der Spielsitzung wurde die Aufnahme gestoppt.

Es folgte eine kurze Reflexionsrunde mit offenen Fragen wie: «Wie hat es sich angefühlt?» oder «Was habt ihr bemerkt?». Danach erhielten die Proband:innen den Fragebogen zur Selbsteinschätzung, der erklärt und später gemeinsam am Tisch ausgefüllt wurde. Dabei konnten Rückfragen gestellt und Unklarheiten geklärt werden. Zum Abschluss bedankte sich der Versuchsleiter, beantwortete letzte Fragen und überreichte LevelHub-Sticker als Dank für die Teilnahme.



Abbildung 10. Station der Datenerhebung mit eingerichteter Konsole, Videokamera und angeschriebenen Sitzplätzen (eigenes Foto).

Die Erhebung erfolgte in einem eigens dafür eingerichteten Raum mit vorbereiteter Station (siehe Abbildung 10). Den Proband:innen wurden nummerierte Sitzplätze (A-D) zugewiesen, um eine anonyme Zuordnung zwischen Selbsteinschätzungen und Videodaten zu ermöglichen. Die ebenfalls zu der Datenerhebung gehörende videogestützte Beobachtungsanalyse wird im nächsten Kapitel ausführlicher beschrieben.

3.2.3 Videogestützte Beobachtungsanalyse

Zur Erfassung der kooperativen Verhaltensweisen im Spiel wurde eine videogestützte Beobachtungsanalyse durchgeführt. Dieses Verfahren ermöglichte eine differenzierte und rekonstruierbare Bewertung des Gruppengeschehens im Nachgang der Spielsitzungen und bildet eine zentrale Datengrundlage der vorliegenden Untersuchung.

Beobachtung von Verhaltensweisen

Die Beobachtung stellt eine zentrale Methode zur Erfassung sozialer und interaktiver Prozesse dar. In Abgrenzung zur alltäglichen Beobachtung ist wissenschaftliche Beobachtung durch einen systematischen, regelgeleiteten und dokumentierten Ablauf charakterisiert (Bortz & Döring, 2006). Sie zielt auf die Erhebung ausgewählter, psychologisch relevanter Verhaltensweisen und erfordert einen zuvor definierten Beobachtungsplan. Darin sollte festgelegt sein, was beobachtet wird, wann, wo und wie lange die Beobachtung stattfindet und wie mit Interpretationsspielräumen umzugehen ist (Hussy, Schreier & Echterhoff, 2013).

Gleichzeitig ist die Beobachtung kognitiv anspruchsvoll: Beobachtende müssen in Echtzeit eine Vielzahl paralleler Verhaltensprozesse wahrnehmen, gewichten und einordnen – insbesondere, wenn mehrere Personen gleichzeitig interagieren. Dabei kann es zu Verzerrungen durch persönliche Erfahrungen und Erwartungen kommen (Higgins & Bargh, 1987; Kleinmann, 2016). Die visuelle Wahrnehmung ist stets selektiv, sodass eine vollständige und objektive Erfassung kaum möglich ist (Kleinmann, 2016).

Einsatz videogestützter Verfahren

Um diesen Herausforderungen zu begegnen, werden in der psychologischen Forschung vermehrt videogestützte Beobachtungsanalysen eingesetzt. Sie ermöglichen es, komplexe Situationen mehrfach, aus unterschiedlichen Perspektiven und unter kontrollierten Bedingungen auszuwerten. Wie verschiedene Studien belegen, eignet sich die Videoanalyse zur Evaluation von Kompetenzen – sowohl im medizinisch-pädagogischen Kontext als auch in simulationsbasierten Lernumgebungen (Brimble, 2008; Campbell, Howie & Murray, 1995; Hays, Jones, Adkins & McKain, 1990). Die videogestützte Beobachtungsanalyse erfordert jedoch gute Planung, Schulung und Instrumentenwahl, um reliable Ergebnisse zu liefern (Haidet, Tate, Divirgilio-Thomas, Kolanowski & Happ, 2009).

In der vorliegenden Untersuchung wurde die videogestützte Beobachtungsanalyse gewählt, um der komplexen Dynamik kooperativer Spielsituationen gerecht zu werden. Sie erlaubt eine objektivere und rekonstruierbare Bewertung von Verhaltenssequenzen und stellt eine praktikable Methode zur systematischen Erfassung von Interaktionen dar.

Die videogestützte Datenerhebung erfolgte mithilfe einer 360°-Kamera, die so positioniert wurde, dass sowohl die Proband:innen als auch der Bildschirm erfasst wurden. Die Aufzeichnung startete mit Spielbeginn und lief während der gesamten Spieldauer sowie der zwischengelagerten Gruppendiskussion. Dadurch konnte der gesamte Kommunikations- und Entscheidungsprozess innerhalb des Spiels festgehalten werden.

Für die nachträgliche Analyse wurde das Programm Insta360 Studio (Arashi Vision Inc., 2023) verwendet, welches eine flexible Navigation innerhalb der 360°-Perspektive erlaubt. Dies ermöglichte es den Bewertenden, situationsspezifisch zwischen Bildschirm- und Gruppeninteraktionen zu wechseln. Dies war insbesondere für die Bewertung komplexer Teamprozesse entscheidend, da sowohl die Spielhandlungen als auch die dazugehörige Kommunikation simultan nachvollzogen werden konnten.

Bewertung mittels VAS-KOOP

Die Fremdeinschätzung der beobachteten Verhaltensweisen erfolgte durch zwei unabhängige Personen aus dem LevelHub-Team. Eine der Bewertenden verfügte über Vorerfahrung in der Verhaltensbeobachtung, während die zweite Person neu im Team war und bisher kaum entsprechende Erfahrungen mitbrachte. Letztere wurde vom Versuchsleiter ausführlich in die Bewertungsdimensionen, Skalen und das Vorgehen eingeführt.

Die aufgezeichneten Videosequenzen wurden mithilfe des Verhaltensankersystems VAS-KOOP für Fremdeinschätzungen (Anhang D) ausgewertet. Zur Unterstützung wurde das ursprünglich für die Selbsteinschätzung konzipierte VAS-KOOP-Instrument in ein Excel-Dokument überführt und um Definitionen sowie Beschreibungen der einzelnen Dimensionen ergänzt. Dies diente als Grundlage für die strukturierte Bewertung.

Während der Bewertung konnten die Videos beliebig pausiert, zurückgespult oder erneut abgespielt werden. Die Beurteilungen wurden systematisch in eine Matrix (Anhang H) eingetragen, die eine eindeutige Zuordnung nach Gruppe, Sitzplatz und Bewertungsdimension

ermöglichte. Nicht belegte Sitzplätze wurden mit einem «X» markiert. Abschliessend wurden die beiden ausgefüllten Bewertungsmatrizen zusammengeführt und für die statistische Analyse vorbereitet.

Das angepasste Beobachtungsinstrument VAS-KOOP bildet somit die konzeptionelle Grundlage der vorliegenden Untersuchung. Das folgende Kapitel zeigt auf, welche Evaluationskriterien statistisch geprüft wurden, um die Qualität und Aussagekraft des entwickelten Instruments systematisch zu bewerten.

3.3 Datenauswertung

Zur Überprüfung der Anwendbarkeit des entwickelten Verhaltensankersystems VAS-KOOP wurden drei zentrale Evaluationskriterien herangezogen: interne Konsistenz, Interrater-Reliabilität sowie die Konstruktvalidität durch Selbst- und Fremdeinschätzung (siehe Tabelle 2). Diese Auswahl orientiert sich an ähnlichen Evaluationsstudien (Arora et al., 2011; Mitchell et al., 2012; Yule et al., 2018), die ein systematisches Vorgehen zur Prüfung von Verhaltensankersystemen beschreiben.

Tabelle 2

Zentrale Evaluationskriterien der Untersuchung mit Bewertungsfragen und Datenquellen

	Test	Evaluationsfrage	Datenquelle
Reliabilität	Interne Konsistenz	Stimmen die Bewertungen auf Elementebene mit den zugehörigen Kategoriebewertungen überein?	Berechnung der mittleren absoluten Differenz zwischen den Bewertungen der Einzelelemente und der jeweiligen Kategorie.
	Interrater-Reliabilität	Inwieweit stimmen die Beurteilungen zweier unabhängiger Bewertenden bei der Anwendung des VAS-KOOP überein?	Berechnung des Pearson-Korrelationskoeffizienten zur Erfassung des Zusammenhangs zwischen beiden Bewertenden.
Validität	Konstruktvalidität	In welchem Ausmass stimmen die Selbstbewertungen der Proband:innen mit den Fremdeinschätzungen durch die Bewertende überein?	Berechnung des Pearson-Korrelationskoeffizienten zur Analyse des Zusammenhangs zwischen Selbst- und Fremdeinschätzung.

Zur Überprüfung der internen Konsistenz des VAS-KOOP wurde Cronbachs α berechnet, basierend auf den drei Elementbewertungen innerhalb jeder Kategorie. Werte im Bereich von 0.71 bis 0.80 gelten gemäss etablierter Empfehlungen als ausreichend zuverlässig, um den Anforderungen an Validität und Messgenauigkeit gerecht zu werden (Yule et al., 2018).

Die Interrater-Reliabilität wurde berechnet, um die Übereinstimmung zwischen den Bewertungen der beiden unabhängigen Beobachtenden zu quantifizieren. Sie spielt besonders für Verhaltensbeobachtungen eine zentrale Rolle (Bakeman & Gottman, 1997). Hierzu wurde der Übereinstimmungskoeffizienten r_{wg} berechnet (James, Demaree & Wolf, 1993). Dieses Verfahren wurde gewählt, da es bereits in einer Evaluationsstudie (Mitchell et al., 2012) eingesetzt wurde. Darin wurde ein Wert von > 0.7 als akzeptabler Grad der Übereinstimmung definiert.

Zur Überprüfung der Konstruktvalidität wurde der Pearson-Korrelationskoeffizient zwischen der Fremdeinschätzung durch die Bewertenden und der Selbsteinschätzung der Proband:innen berechnet. Ein solcher Vergleich ist zulässig, sofern beide Erhebungsverfahren dasselbe theoretische Konstrukt abbilden (Schmidt-Atzert, Krumm & Amelang, 2021). Ziel dieses Vorgehens war es, neben der Fremdeinschätzung einen weiteren Bezugspunkt einzubeziehen und so eine fundierte Grundlage für weiterführende Reflexionen zu schaffen. Methodisch orientiert sich der Ansatz an Arora et al. (2011), die einen ähnlichen Vergleich im Kontext nicht-technischer Fähigkeiten vorgenommen haben.

Für alle Bewertungen kamen deskriptive Statistiken (Mittelwerte, Standardabweichungen) sowie Korrelationsanalysen zum Einsatz. Die Datenanalyse wurde mit der Statistiksoftware Jamovi (The jamovi project, 2025) durchgeführt.

Zusammenfassend erfolgte die Datenauswertung entlang etablierter Gütekriterien, um die Aussagekraft und Anwendbarkeit des VAS-KOOP systematisch zu prüfen. Das folgende Kapitel präsentiert die konkreten Ergebnisse dieser Analysen.

4 Ergebnisse

Das folgende Kapitel stellt die Ergebnisse der statistischen Auswertung vor. Ziel war es, die Qualität und Aussagekraft des entwickelten Beobachtungsinstruments VAS-KOOP zu prüfen. Dazu wurden zentrale Gütekriterien wie interne Konsistenz (4.2), Interrater-Reliabilität (4.3) sowie die Konstruktvalidität anhand von Selbst- und Fremdeinschätzungen (4.4) untersucht.

4.1 Deskriptiv Analyse

Für die statistische Analyse wurden sämtliche erhobene Daten berücksichtigt. Diese wurden in Jamovi (The jamovi project, 2025) integriert und den Sitzplätzen entsprechend dem Spielzeitpunkt systematisch zugeordnet (z. B. 1A, 1B, 2A, 2B). Insgesamt wurden 27 Variablen erfasst, verteilt auf drei Perspektiven (Bewertung A, Bewertung B und Selbsteinschätzung) sowie neun Dimensionen, die den drei Hauptkategorien Entscheidungsfindung, Kommunikation & Teamarbeit und Führung zugeordnet sind.

Die Benennung der Variablen folgte einem festen Schema (z. B. e1A: Entscheidungsfindung, Unterdimension 1, Bewertung A). Zur Berechnung der Korrelationskoeffizienten wurden anschliessend für jede Kategorie Mittelwerte gebildet, wodurch neun Variablen (z. B. MEAN_eA, MEAN_kA) entstanden. Sämtliche Items weisen metrische Werte auf. Die niedrigste Ausprägung (1) und die höchste Ausprägung (5) entsprechen jeweils den *negative* und *positiven* Verhaltensindikatoren pro Dimension.

Die Abbildung 11 zeigt die Mittelwerte der drei Kategorien Entscheidungsfindung, Kommunikation & Teamarbeit sowie Führung aus der Sicht von Bewertung A, Bewertung B und der Selbsteinschätzung. Die Werte liegen zwischen 3.79 und 4. Auffällig ist, dass die Selbsteinschätzungen in allen Kategorien leicht über den Fremdeinschätzungen liegen – besonders deutlich in den Kategorien Entscheidungsfindung sowie Kommunikation & Teamarbeit. Zudem weist Bewertung B durchgehend die niedrigsten Werte auf.

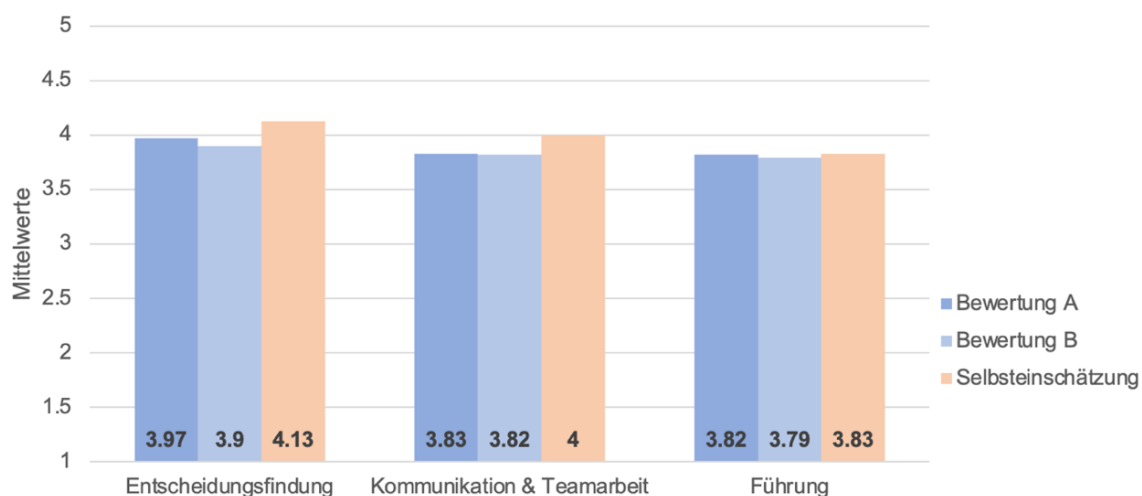


Abbildung 11. Mittelwerte der gemessenen Kategorien: Entscheidungsfindung, Kommunikation & Teamarbeit sowie Führung über die drei Perspektiven hinweg; N = 30.

Die Tabelle 3 zeigt die Mittelwerte und Standardabweichungen für alle Kategorien sowie deren untergeordnete Dimensionen aus den Perspektiven von Bewertung A, Bewertung B und der Selbsteinschätzung. In nahezu allen Fällen liegen die Werte der Selbsteinschätzung leicht über denen der beiden Fremdeinschätzungen, besonders deutlich bei *Betrachten von Optionen* ($M = 4.30$, $SD = 0.70$) und *Informationsaustausch* ($M = 4.20$, $SD = 0.81$). Bewertung B weist mehrfach die niedrigsten Mittelwerte auf, etwa bei *Ermutigung anderer* ($M = 3.53$, $SD = 0.82$). Die geringsten Abweichungen zwischen den drei Perspektiven zeigen sich in der Kategorie *Führung*. Die Standardabweichungen der drei Kategorien reichen von 0.55 bis 0.86, wobei die Selbsteinschätzungen die geringste Streuungen aufweisen.

Tabelle 3

Darstellung der Mittelwerte und Standardabweichungen sämtlicher gemessenen Dimensionen und Kategorien über die drei Perspektiven hinweg

	Bewertung A		Bewertung B		Selbsteinschätzung	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Entscheidungsfindung	3.97	0.69	3.90	0.74	4.13	0.60
Betrachten von Optionen	3.97	0.85	3.80	0.89	4.30	0.70
Auswahl und Kommunikation einer Option	4.00	0.83	3.87	0.78	3.87	0.78
Umsetzung & Überprüfung von Entscheidungen	3.93	0.83	4.03	0.85	4.23	0.86
Kommunikation & Teamarbeit	3.83	0.69	3.82	0.86	4.00	0.55
Informationsaustausch	3.90	0.80	3.93	0.87	4.20	0.81
Gemeinsames Verständnis herstellen	3.73	0.91	3.83	0.99	4.00	0.83
Koordinierung von Teamaktivitäten	3.87	0.82	3.70	0.99	3.80	0.66
Führung	3.82	0.73	3.79	0.78	3.83	0.70
Ermutigung anderer	3.73	0.79	3.53	0.82	3.63	0.89
Unterstützung anderer	3.80	0.89	3.97	0.89	3.90	0.92
Umgang mit Druck	3.93	0.87	3.87	0.90	3.97	0.93

Anmerkungen. *M* für Mittelwerte; *SD* für Standardabweichungen; $N = 30$.

Zur Prüfung der Normalverteilung wurde der Shapiro-Wilk-Test angewendet. Sechs der neun untersuchten Variablen zeigen keine signifikante Abweichung und gelten damit als normalverteilt. Drei Variablen – die Selbsteinschätzung der Entscheidungsfindung, die Bewertung B in der Kategorie Kommunikation & Teamarbeit sowie die Bewertung A in der Kategorie der Entscheidungsfindung – weichen signifikant ab und sind nicht normalverteilt. Diese Ergebnisse werden bei der Interpretation entsprechend vorsichtig berücksichtigt.

4.2 Interne Konsistenz

Die interne Konsistenz des VAS-KOOP-Instruments wurde über Cronbachs α für die drei Hauptkategorien und deren Items berechnet – jeweils getrennt für Bewertung A, Bewertung B und die Selbsteinschätzung (siehe Tabelle 4).

Tabelle 4

Reliabilität des VAS-KOOP-Instruments zur Bewertung überfachlicher Fähigkeiten in drei Kategorien

	Bewertung A		Bewertung B		Selbsteinschätzung	
	α^a	α^b	α^a	α^b	α^a	α^b
Entscheidungsfindung	0.77		0.86		0.65	
Betrachten von Optionen		0.75		0.81		0.41
Auswahl und Kommunikation einer Option		0.70		0.79		0.43
Umsetzung & Überprüfung von Entscheidungen		0.61		0.79		0.78
Kommunikation & Teamarbeit	0.76		0.89		0.51	
Informationsaustausch		0.63		0.89		0.47
Gemeinsames Verständnis herstellen		0.71		0.79		0.34
Kordinierung von Teamaktivitäten		0.69		0.86		0.41
Führung	0.82		0.88		0.66	
Ermutigung anderer		0.82		0.78		0.61
Unterstützung anderer		0.73		0.86		0.38
Umgang mit Druck		0.72		0.85		0.67

Anmerkungen. ^a α für die jeweilige VAS-KOOP-Kategorie, basierend auf den drei zugehörigen Elementen; ^b α für die Kategorie, sofern das jeweilige Element entfernt würde; N = 30.

Die α -Werte für die Bewertungen A und B lagen in allen drei Kategorien im Bereich zwischen 0.76 und 0.89, was auf eine insgesamt stabile interne Konsistenz hinweist. Die zugehörigen Elemente zeigten ebenfalls durchweg α -Werte im mittleren bis hohen Bereich. Die Gesamtreliabilität des VAS-KOOP über alle Kategorien hinweg betrug für Bewertung A $\alpha = 0.90$ und für Bewertung B $\alpha = 0.96$.

Bei den Selbsteinschätzungen fielen die Werte für die interne Konsistenz insgesamt niedriger aus. Die berechneten α -Werte auf Kategorieebene lagen zwischen 0.51 und 0.66. Einzelanalysen der Items zeigten jedoch, dass durch die Entfernung bestimmter Elemente eine Verbesserung der internen Konsistenz möglich wäre. So würde beispielsweise der α -Wert der Kategorie Entscheidungsfindung von 0.65 auf 0.78 ansteigen, wenn die Dimension *Umsetzung & Überprüfung von Entscheidungen* entfernt würde. Die Gesamtreliabilität über alle Kategorien hinweg betrug für die Selbsteinschätzung $\alpha = 0.84$.

4.3 Interrater-Reliabilität

Mit der Berechnung der Interrater-Reliabilität sollte unter anderem die Annahme geprüft werden, ob die Beurteilungen zweier unabhängigen Bewertenden miteinander übereinstimmen. Dazu wurde der r_{wg} als Mass für die Within-Group Agreement (James et al., 1993) bzw. der Pearson-Korrelationskoeffizient berechnet. Die Analyse umfasste drei Vergleichspaare: Bewertung A und Bewertung B, Bewertung A und Selbsteinschätzung sowie Bewertung B und Selbsteinschätzung. Die Ergebnisse dieser Gruppenvergleiche sind in Tabelle 5 dargestellt.

Tabelle 5

Übereinstimmung der erhobenen Kategorien innerhalb der Gruppe (r_{wg}) über drei Vergleichspaare hinweg

	Bewertung A Bewertung B	Bewertung A Selbsteinschätzung	Bewertung B Selbsteinschätzung
	r_{wg}	r_{wg}	r_{wg}
Entscheidungsfindung	0.65***	0.60***	0.12
Kommunikation & Teamarbeit	0.69***	0.29	0.29
Führung	0.70***	0.34	0.02

Anmerkungen. *** $p < .001$; r_{wg} : Pearson-Korrelation für Within-Group Agreement; N = 30.

Für die Fremdeinschätzungen von Person A und Person B zeigten alle drei Kategorien signifikante hohe Übereinstimmungen:

- Entscheidungsfindung: $r_{wg} = .65$, $p < .001$, 90% KI für r_{wg} [0.42, 0.80]
- Kommunikation & Teamarbeit: $r_{wg} = .69$, $p < .001$, 90% KI für r_{wg} [0.49, 0.82]
- Führung: $r_{wg} = .70$, $p < .001$, 90% KI für r_{wg} [0.50, 0.83]

Die Abbildung 12 illustriert die Einzelergebnisse der beiden Bewertenden für alle der 30 Proband:innen, getrennt nach den drei. Jede Linie repräsentiert eine Bewertungsperspektive in einer spezifischen Kategorie. Die Verläufe der sechs Kurven zeigen weitgehend Ähnlichkeiten mit wenigen sichtbaren Abweichungen bei einzelnen Fällen.

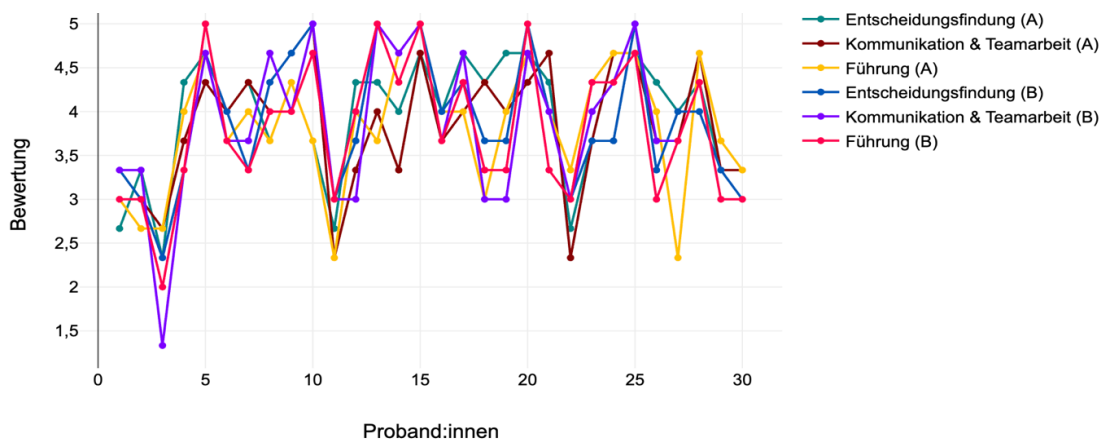


Abbildung 12. Mittelwerte der Bewertungen A und B getrennt nach den Kategorien Entscheidungsfindung, Kommunikation & Teamarbeit und Führung; N = 30 (erstellt mit datatab.de).

4.4 Vergleich Fremd- und Selbsteinschätzung

In einem weiteren Schritt wurde überprüft, inwieweit die Selbsteinschätzungen der Proband:innen mit den Fremdeinschätzungen übereinstimmen. Ziel war es, die Annahme zu prüfen, ob sie Fremd- und Selbsteinschätzungen miteinander in Verbindung stehen.

Dazu wurde r_{wg} als Mass der Übereinstimmung zwischen den jeweiligen Bewertungen berechnet. Untersucht wurden mithilfe des Pearson-Korrelationskoeffizienten die Zusammenhänge zwischen der Selbsteinschätzung und den Fremdeinschätzungen durch Person A bzw. Person B über alle drei Hauptkategorien hinweg. Tabelle 6 gibt einen detaillierten Überblick zu den Übereinstimmungswerten der jeweiligen Vergleichspaare.

Zwischen Bewertung A und der Selbsteinschätzung zeigte sich lediglich in der Kategorie Entscheidungsfindung ein signifikanter Zusammenhang:

- Entscheidungsfindung: $r_{wg} = .60$, $p < .001$, 90% KI für r_{wg} [0.35, 0.76]
- Kommunikation & Teamarbeit: $r_{wg} = .29$, $p = .116$
- Führung: $r_{wg} = .34$, $p = .07$

Die Korrelationen zwischen Bewertung B und der Selbsteinschätzung fielen geringer aus und bewegten sich ebenfalls im nicht signifikanten Bereich:

- Entscheidungsfindung: $r_{wg} = .12$, $p = .535$
- Kommunikation & Teamarbeit: $r_{wg} = .29$, $p = .127$
- Führung: $r_{wg} = .02$, $p = .898$.

4.5 Zusammenfassung der Ergebnisse

Die interne Konsistenz des entwickelten Instruments VAS-KOOP wurde mithilfe von Cronbachs α für drei Hauptkategorien sowie deren zugehörige Items analysiert. Die Ergebnisse zeigten für die beiden Fremdeinschätzungen α -Werte zwischen 0.76 und 0.89, was auf eine durchweg stabile Reliabilität hinweist. Die Selbsteinschätzungen ergaben geringere α -Werte im Bereich von 0.51 bis 0.66. Die Gesamtreliabilität lag bei $\alpha = 0.90$ für Bewertung A, bei $\alpha = 0.96$ für Bewertung B und bei $\alpha = 0.84$ für die Selbsteinschätzung.

Zur Erfassung der Interrater-Reliabilität wurde die Übereinstimmung der Beurteilungen zweier unabhängiger Bewertender analysiert. Alle drei Kategorien erreichten dabei signifikante Werte zwischen r_{wg} 0.65 und 0.70. Diese Ergebnisse deuten auf eine hohe Übereinstimmung zwischen den beiden Bewertenden hin.

Die Übereinstimmung zwischen Selbst- und Fremdeinschätzungen fiel deutlich schwächer aus. Nur in der Kategorie Entscheidungsfindung wurde in einer Kombination zwischen Bewertung A und Selbsteinschätzung ein signifikanter Zusammenhang ermittelt ($r_{wg} = .60$, $p < .001$, 90% KI für r_{wg} [0.35, 0.76]). Aufgrund der nicht vorhandenen Normalverteilung dieser Kategorien gilt es, diese Korrelation jedoch mit Vorsicht zu interpretieren. In den übrigen Kategorien blieben die Werte unterhalb der Signifikanzgrenze.

5 Diskussion

Die Ergebnisse der vorliegenden Untersuchung liefern erste Hinweise darauf, dass überfachliche Fähigkeiten in kooperativen Spielsituationen strukturiert beobachtet und bewertet werden können – ein zentrales Anliegen, das bereits in der Ausgangslage dieser Arbeit formuliert wurde. Vor dem Hintergrund der zunehmenden Relevanz überfachlicher Kompetenzen in Bildung, Arbeitswelt und Gesellschaft wurde mit dem VAS-KOOP ein Instrument entwickelt, das diesen Anforderungen auf innovative Weise begegnet. Die Kombination aus spielerischem Setting und standardisierter Verhaltensbeobachtung schliesst dabei eine bislang wenig erforschte Lücke in der diagnostischen Praxis mit GBAs. Das folgende Kapitel diskutiert die zentralen Ergebnisse vor diesem Hintergrund, reflektiert methodische Aspekte kritisch und leitet Implikationen für Forschung und Anwendung ab.

5.1 Zusammenfassung der Arbeit

Ausgangspunkt dieser Arbeit war die bestehende Herausforderung, kooperatives Spielverhalten im Rahmen der LevelHub-Methodik systematisch und objektiv zu erfassen. Bisher beruhte die Bewertung von LevelHub primär auf subjektiven Beobachtungen, was die Vergleichbarkeit erschwerte und die Reliabilität einschränkte. Ziel der vorliegenden Untersuchung war daher die Entwicklung und Überprüfung eines Verhaltensankersystems zur systematischen Beurteilung überfachlicher Fähigkeiten in kooperativen Spielsituationen.

Der theoretische Rahmen der Arbeit stützte sich auf drei zentrale Bausteine:

1. Die LevelHub-Methodik, die überfachliche Kompetenzen wie Kollaboration in spielerischen Gruppensettings beobachtbar macht.
2. Das Konzept des GBA, das interaktive Spielsituationen zur Erfassung von beobachtbarem Verhalten nutzt.
3. Strukturierte Verhaltensankersysteme wie NOTSS, das auf Basis konkreter Verhaltensindikatoren eine objektive Einschätzung überfachlicher Fähigkeiten erlaubt und für diese Arbeit adaptiert wurde.

Das Instrument VAS-KOOP wurde in einem mehrstufigen Vorgehen entwickelt: Zunächst wurden durch eine Recherche und ein Experteninterview geeignete Modelle identifiziert. Anschliessend erfolgte die Modifikation des bestehenden Modells, wobei die CIT zur Entwicklung kontextspezifischer Verhaltensindikatoren eingesetzt wurde. Das finale System umfasst neun Items, verteilt auf die drei Hauptkategorien Entscheidungsfindung, Kommunikation & Teamarbeit und Führung. Es wurde sowohl für die Selbst- als auch für die Fremdeinschätzung operationalisiert.

Die empirische Überprüfung erfolgte im Rahmen einer Datenerhebung mit insgesamt 30 Proband:innen. Die Beobachtungen wurden videogestützt dokumentiert und retrospektiv durch zwei unabhängige Bewertende eingeschätzt. Ergänzend füllten die Proband:innen nach den Gaming-Sessions eine Selbsteinschätzung aus.

Die Ergebnisse zeigen eine gute interne Konsistenz des Instruments und eine hohe Übereinstimmung zwischen den beiden Bewertenden (Interrater-Reliabilität). Für die Konstruktvalidität ergaben sich nur schwache Hinweise. Lediglich in der Kategorie Entscheidungsfindung bei einer der Bewertenden konnte ein signifikanter Zusammenhang mit der Selbsteinschätzung festgestellt werden, was jedoch mit Vorsicht zu bewerten ist.

5.2 Interpretation der Ergebnisse

Die vorliegenden Ergebnisse zeigen, dass das entwickelte Verhaltensankersystem VAS-KOOP erfolgreich auf kooperative Videospielsituationen übertragen werden konnte. Die zentrale Forschungsfrage – ob ein bestehendes System wie NOTSS (Yule, Flin, Paterson-Brown, Maran, et al., 2006) auf ein Spielsetting anwendbar ist – lässt sich grundsätzlich bejahen. Mithilfe etablierter methodischer Verfahren konnte ein reliables Instrument zur Bewertung überfachlicher Fähigkeiten im Spielkontext entwickelt werden.

Das VAS-KOOP erwies sich dabei als praktikables System zur Einschätzung von überfachlichen Fähigkeiten wie Entscheidungsfindung, Teamarbeit und Führung. Besonders hervorzuheben ist die hohe Übereinstimmung beider unabhängigen Bewertenden, die auf eine gute Objektivierbarkeit des Instruments hinweist. Gleichzeitig zeigen die Ergebnisse niedrige Zusammenhänge zwischen Selbst- und Fremdeinschätzungen. Hier ergeben sich jedoch neue Chancen: Die beobachteten Diskrepanzen liefern Hinweise auf blinde Flecken und damit wertvolle Anknüpfungspunkte für Entwicklungs- und Reflexionsprozesse. In Summe lässt sich festhalten, dass das VAS-KOOP erfolgreich auf den spielerischen, kooperativen Kontext übertragen wurde und eine solide Grundlage für weitere Forschung sowie praktische Anwendungen bietet.

Interne Konsistenz als Qualitätsmerkmal

Die gute interne Konsistenz des VAS-KOOP deutet darauf hin, dass die Items innerhalb der jeweiligen Dimensionen zuverlässig ein gemeinsames Konstrukt erfassen. Eine mögliche Erklärung hierfür liegt in der Orientierung am etablierten NOTSS-System (Yule et al., 2018), das bereits validierte Kategorien nicht-technischer Fähigkeiten bereitstellt. Die klare Gliederung in drei thematisch fokussierte Kategorien – Entscheidungsfindung, Kommunikation & Teamarbeit sowie Führung – stärkt die Messqualität des Instruments. Die Übernahme und Anpassung dieser Struktur unterstützten die Entwicklung inhaltlich konsistenter Items. Verhaltensindikatoren mit Fokus auf weniger gleichzeitig zu beobachtende Aspekte könnten zu einer besseren Orientierung führen, Interpretationsspielräume reduzieren und die Reliabilität zusätzlich erhöhen (Kleinmann, 2016; Obermann, 2002).

Objektivität durch Interrater-Reliabilität

Die Analyse der Interrater-Reliabilität zeigt, dass die Beurteilungen der beiden unabhängigen Bewertenden in einem statistisch signifikanten Zusammenhang stehen. Für die Dimensionen Entscheidungsfindung, Kommunikation & Teamarbeit sowie Führung wurden jeweils signifikante Korrelationen festgestellt, die auf eine weitgehende

Übereinstimmung in den Beurteilungen hinweisen. Auch wenn die Werte knapp unter der von Yule et al. (2018) empfohlenen Schwelle für Zuverlässigkeit liegen, sprechen sie dennoch für eine gute Anwendbarkeit und Verständlichkeit des Verhaltensankersystems. Eine verlässliche Interrater-Reliabilität ist wesentlich, um die Objektivität der Beobachtungsdaten zu sichern und deren Aussagekraft zu gewährleisten (Tinsley & Weiss, 1975), was durch diese Ergebnisse grundsätzlich gestützt wird.

Diskrepanz zwischen Selbst- und Fremdwahrnehmung

Die Übereinstimmung zwischen Selbst- und Fremdeinschätzungen im VAS-KOOP fiel insgesamt gering aus. Ein signifikanter Zusammenhang wurde lediglich in der Dimension Entscheidungsfindung mit einer der beiden Bewertenden festgestellt. Die Konstruktvalidität des Instruments fällt somit ungenügend aus. In den meisten Fällen bildete das VAS-KOOP in der Fremdeinschätzung nicht das gleiche Bild wie in der Selbsteinschätzung ab. Solche Diskrepanzen sind aus der Forschung bekannt und lassen sich dadurch erklären, dass Selbstbeurteilungen auf subjektive, nicht sichtbare Informationen wie Gedanken, Absichten oder Gefühle basieren (Falchikov & Boud, 1989). Studien zeigen daher häufig geringe Übereinstimmungen (Bieri Buschor & Schuler Braunschweig, 2011; Clausen, Möller & Strauß, 2000).

Während Selbsteinschätzungen eine relevante prognostische Funktion für individuelle Entwicklungsprozesse haben (Braun, Sheikh & Hannover, 2011), gilt beobachtbares Verhalten als verlässlicher Prädiktor für zukünftiges Verhalten (Ouellette & Wood, 1998; Wernimont & Campbell, 1968). Die Diskrepanz zwischen beiden Perspektiven sollte daher nicht nur als Schwäche gewertet werden. Sie kann vielmehr als Ausgangspunkt für gezielte Reflexions- und Lernprozesse genutzt werden.

Blinde Flecken und verzerrte Selbstwahrnehmung

Ein bewährtes Modell zur Veranschaulichung der Unterschiede zwischen Selbst- und Fremdwahrnehmung ist das Johari-Fenster (Luft & Ingham, 1955). Es beschreibt den sogenannten *blinden Fleck* – jene Persönlichkeits- oder Verhaltensaspekte, die von aussen wahrgenommen werden, einem selbst jedoch nicht bewusst sind. Durch strukturierte Fremdeinschätzungen kann dieser Bereich sichtbar gemacht werden – ein Prozess, der gerade im Hinblick auf die Entwicklung überfachlicher Fähigkeiten zentral ist.

In der vorliegenden Erhebung fielen die Selbsteinschätzungen insgesamt höher aus als die Fremdeinschätzungen. Dies könnte – im Sinne des Dunning-Kruger-Effekts (Kruger & Dunning, 1999) – darauf hinweisen, dass Personen mit geringerer Ausprägung in bestimmten überfachlichen Fähigkeiten ihre Kompetenzen überschätzen. Besonders relevant ist dies im Kontext arbeitsintegrativer Massnahmen, aus denen ein Grossteil der Stichprobe stammt. In solchen Settings sind Themen wie Selbstbild, Selbstwirksamkeit und soziale Vergleichsprozesse besonders präsent. Forschende verweisen darauf, dass durch biografische Brüche, Misserfolgserfahrungen oder fehlende Rückmeldestrukturen verzerrte Selbstwahrnehmungen begünstigt werden können (Dewe, 2009; Galuske & Müller, 2002). Diese Beobachtung wird durch Einschätzungen von Abteilungsleitenden der

Integrationsmassnahmen im Netzwerk Grenchen gestützt. Gleichzeitig könnten die retrospektive Fremdeinschätzungen anhand der Videoaufnahmen zu kritischeren Urteilen geführt haben, da Situationen mehrfach betrachtet und genauer analysiert werden konnten.

Die Diskrepanz zwischen Selbst- und Fremdwahrnehmung war somit nicht nur erwartbar, sondern bietet einen wertvollen Ansatzpunkt für gezielte Reflexions- und Entwicklungsprozesse. Das VAS-KOOP kann in diesem Sinne zur Sichtbarmachung blinder Flecken beitragen und einen Impuls für Veränderungsprozesse setzen.

Erklärungen zur Übereinstimmung in der Entscheidungsfindung

Da sowohl die Selbsteinschätzung als auch die Fremdeinschätzung in der Kategorie Entscheidungsfindung signifikant von der Normalverteilung abweichen, sind die Ergebnisse mit Vorsicht zu interpretieren. Die auffällig hohe Übereinstimmung zwischen Fremd- und Selbsteinschätzung in dieser Kategorie könnte teilweise durch das Setting begünstigt worden sein. Das kooperative Videospiel sowie die anschliessende Gruppendiskussion fordern explizite Entscheidungen und machen entsprechendes Verhalten besonders sichtbar. Entscheidungsverhalten äussert sich oft in konkreten Handlungen oder verbalen Äusserungen, was die Beobachtung möglicherweise erleichtert (Salas, Shuffler, Thayer, Bedwell & Lazzara, 2015). Im Gegensatz dazu beinhalten die Dimensionen Kommunikation & Teamarbeit sowie Führung komplexere relationale Aspekte, die stärker im Zusammenspiel mit anderen Personen stehen und dadurch eine isolierte Beobachtung erschweren. Zudem ist zu berücksichtigen, dass der signifikante Zusammenhang nur bei einer der beiden bewertenden Personen festgestellt wurde, was die Aussagekraft zusätzlich relativiert.

Qualitätssicherung durch Schulung der Bewertenden

Die Analyse der Fremdeinschätzungen zeigte Unterschiede zwischen den beiden Bewertenden. Person A, die mit der Entwicklung des Verhaltensankersystem vertraut war, erzielte höhere Übereinstimmungen mit den Selbsteinschätzungen als Person B. Letztere erhielt vor der Bewertung lediglich eine kurze Einführung in die Messung mit VAS-KOOP. Ihre Einschätzungen stimmten weniger mit den Selbsteinschätzungen überein, besonders in den Dimensionen Führung. Obwohl die Korrelationen nicht signifikant sind, deuten die Unterschiede darauf hin, dass Vorbereitung und Schulung die Qualität der Beobachtungsdaten massgeblich beeinflussen können.

Um die Objektivität und Reliabilität des Instruments zu sichern, sind gezielte Trainings notwendig. Obermann (2002) empfiehlt Trainingsformate, die sowohl Selbstreflexion als auch gemeinsame Definitionsarbeit an Bewertungsdimensionen beinhalten. Dies unterstützt Beobachtende dabei, Wahrnehmungsverzerrungen zu erkennen und ein gemeinsames Verständnis der Bewertungsstandards zu entwickeln. Kleinmann (2016) schlägt eine Kombination aus Verhaltensbeobachtungstraining und Bezugsrahmentraining vor. Während ersteres die Trennung von Beobachtung und Bewertung fördert, veranschaulicht zweiteres durch realistische Verhaltensbeispiele die Bandbreite möglicher Ausprägungen. Beide Trainingsformen zusammen verbessern nachweislich die Beurteilungsgenauigkeit und reduzieren typische Beobachtungsfehler.

Potenzial als diagnostisches Instrument

Die Ergebnisse deuten darauf hin, dass sich die LevelHub-Methodik grundsätzlich als GBA für Personalbeurteilungen oder Teamentwicklung eignet. Das strukturierte Verhaltensankersystem macht überfachliche Kompetenzen wie Entscheidungsfindung, Teamarbeit oder Führung in realitätsnahen Spielsituationen sichtbar und kann die Testmotivation erhöhen. In diagnostischen Kontexten steht häufig die Bewertung anhand eines festgelegten Kriteriums im Vordergrund, was deutlich höhere Anforderungen an die Validität des Verfahrens stellt. Um diesem Anspruch gerecht zu werden, müsste das Instrument mit etablierten, validierten Verfahren kombiniert und im Hinblick auf Kriteriumsvalidität geprüft werden (Kanning, 2019; Schmidt-Atzert et al., 2021).

Förderung überfachlicher Kompetenzen durch GBL

Überfachliche Kompetenzen wie Kollaboration oder Entscheidungsfindung gelten als zentrale Erfolgsfaktoren vor dem Hintergrund gesellschaftlicher und arbeitsweltbezogener Herausforderungen. Die gezielte Förderung dieser Fähigkeiten mittels GBL bietet einen niederschweligen und motivierenden Ansatz. Studien belegen positive Effekte von Videospiele auf die Entwicklung sozialer und kognitiver Kompetenzen (Bailey et al., 2006; Beigi et al., 2025; Connolly et al., 2012; Granic et al., 2014). Zudem deuten Ergebnisse aus simulationsbasierten Trainings auf einen mittelfristigen Erhalt solcher Kompetenzen hin (Yule et al., 2018). Mit ihrer spielbasierten Methodik leistet LevelHub in diesem Kontext einen innovativen Beitrag zur Entwicklung überfachlicher Fähigkeiten.

Zusammenfassend lässt sich festhalten, dass das entwickelte Instrument VAS-KOOP als praktikables und reliables Instrument zur Bewertung überfachlicher Fähigkeiten in kooperativen Spielsettings eingesetzt werden kann. Die gute interne Konsistenz und Interrater-Reliabilität sprechen für eine solide Messqualität. Gleichzeitig weist die geringe Übereinstimmung zwischen Selbst- und Fremdeinschätzung auf wichtige Entwicklungspotenziale hin – sowohl inhaltlich als auch methodisch. Das Instrument bietet somit nicht nur eine Basis zur systematischen Verhaltensbeobachtung, sondern auch Anknüpfungspunkte für Reflexionsprozesse und weiterführende Anwendungen, etwa im Rahmen von GBA oder Kompetenzförderung mittels GBL.

5.3 Limitationen der Untersuchung

Trotz dieser positiven Befunde sind für die vorliegende Untersuchung verschiedene Limitationen zu berücksichtigen, die sowohl die Aussagekraft der Ergebnisse einschränken als auch Hinweise für zukünftige Optimierungen liefern. Im Folgenden werden methodische, inhaltliche und kontextbezogene Einschränkungen diskutiert.

Begrenzungen in der Entwicklung und Auswahl der Beobachtungskriterien

Die Verhaltensindikatoren zur Entwicklung des Instruments erfolgte überwiegend durch LevelHub-Teammitglieder, die mit der Methodik vertraut waren. Dies birgt das Risiko verzerrter Einschätzungen.

Zudem konzentriert sich das entwickelte Instrument lediglich auf drei überfachliche Fähigkeiten, die wiederum nur mit je drei Unterdimensionen erfasst wurden. Eine zu geringe Anzahl an Items oder Facetten kann die inhaltliche Validität eines Tests gefährden (Moosbrugger & Kelava, 2020). Die eingeschränkte Abbildung der Fähigkeiten ist daher kritisch zu betrachten. Weitere Kompetenzen wie Kreativität, Problemlösungsfähigkeit oder Anpassungsfähigkeit, die laut OECD (2018) und World Economic Forum (2025) zu wichtigen Zukunftskompetenzen zählen, wurden nicht berücksichtigt. Obwohl sie für den Umgang mit einer zunehmend komplexen und dynamischen Arbeitswelt entscheidend sind.

Unschärfe der Indikatoren und Optimierung der Bewertungsskala

Einige Dimensionen vereinten in den Verhaltensindikatoren eine Vielzahl teils heterogener Einzelaspekte, darunter z.B.: «überprüft Aktivitäten», «schafft Übersicht», «priorisiert», «sorgt für Effizienz und zeitliche Abstimmung» und «klärt Missverständnisse». Diese Vielzahl erschwerte eine klare Fokussierung und Gewichtung im Bewertungsprozess.

Weiter wurden nur die Extrempole der Skala verbal beschrieben. Mittlere Ausprägungen blieben unkonkret, was bei den Bewertenden zu Unsicherheiten führen konnte. Zudem handelte es sich nicht um eine sogenannte Behaviorally Anchored Rating Scale (BARS), bei der jeder Skalenwert mit konkreten Verhaltensbeispielen verknüpft ist. Der Einsatz solcher Skalen kann die Objektivität und Vergleichbarkeit der Bewertungen erhöhen (Kleinmann, 2016; Smith & Kendall, 1963). Eine noch strukturiertere Alternative stellen Verhaltens-Checklisten dar, die sich zunächst auf das Vorhandensein klar definierter Verhaltensweisen konzentrieren und daraus eine zusammenfassende Bewertung ableiten. Dadurch lassen sich vorschnelle Globalurteile vermeiden und die Konstruktvalidität erhöhen (Obermann, 2002).

Retrospektive Bewertung und Schulungsbedarf

Die videogestützte, retrospektive Analyse erlaubte eine wiederholte Betrachtung von Sequenzen, wodurch differenzierte Einschätzungen möglich wurden. Bei einer Anwendung in Echtzeit-Settings – etwa direkt während Workshops – ist dies nicht gleichermassen umsetzbar. Die simultane Bewertung von neun Dimensionen bei mehreren Personen erfordert hohe kognitive Ressourcen von Beobachtenden. Für den praktischen Einsatz scheint eine Reduktion auf zentrale Kerndimensionen sowie eine modulare Anwendung des Instruments sinnvoll.

Zusätzlich zeigte sich, dass eine kurze Einführung in die Systematik nicht ausreicht. Eine zuverlässige Anwendung des VAS-KOOP erfordert ein gezieltes Training (Kleinmann, 2016; Obermann, 2002). Für weitere Erhebungen sollte der Fokus stärker auf Instruktionen vor der Erhebung gelegt werden. Dies könnte helfen, Bewertungsstandards zu klären, Beobachtungsfehler zu reduzieren und die Beurteilungsgenauigkeit zu verbessern.

Stichprobeneigenschaften und fehlende Rückmeldung zur Anwendbarkeit

Die Stichprobe bestand aus 30 Personen, überwiegend aus arbeitsintegrativen Massnahmen. Eine breitere Rekrutierung aus unterschiedlichen beruflichen Kontexten wäre für eine grössere Übertragbarkeit der Ergebnisse wünschenswert gewesen. Darüber hinaus wurden keine personenbezogenen Variablen wie Alter, Spielerfahrung oder subjektives Stresslevel erfasst – obwohl diese Faktoren potenziell Einfluss auf das beobachtete Verhalten haben könnten.

Ebenso fehlt eine systematische Erhebung zur Akzeptanz und Nutzbarkeit des VAS-KOOP aus Sicht der Proband:innen. Angaben zur Verständlichkeit der Items oder zur Handhabung hätten wertvolle Hinweise für die nutzerorientierte Weiterentwicklung des Instruments liefern können.

6 Schlussfolgerung

Die vorliegende Arbeit verfolgte das Ziel, ein systematisches Beobachtungsinstrument zur Erfassung überfachlicher Fähigkeiten im Kontext eines kooperativen Videospiele zu entwickeln und zu erproben. Aufbauend auf den Ergebnissen und der Diskussion werden in diesem Kapitel zentrale Implikationen, praktische Einsatzmöglichkeiten sowie Ansatzpunkte für die Weiterentwicklung des VAS-KOOP skizziert.

6.1 Implikationen für LevelHub

Mit dem VAS-KOOP wurde ein praxisnahes Beobachtungsinstrument entwickelt, das die systematische Erfassung überfachlicher Fähigkeiten in kooperativen Spielsituationen ermöglicht. Das Instrument trägt zur Qualitätssicherung der LevelHub-Methodik bei, indem es subjektive Einschätzungen durch klar definierte Verhaltensindikatoren ergänzt und so die Beobachtung standardisiert. Durch die Fokussierung auf ausgewählte Kategorien – wie Entscheidungsfindung, Kommunikation & Teamarbeit sowie Führung – wird die Beobachtbarkeit zentraler Kompetenzen erleichtert und eine Vergleichbarkeit zwischen Teilnehmenden ermöglicht.

Darüber hinaus kann die transparente Anwendung des VAS-KOOP als strukturierte Reflexionshilfe dienen. Die Offenlegung der Verhaltensindikatoren kann die Auseinandersetzung mit dem eigenen Handeln anregen und eine fundierte Basis für Feedback- und Lernprozesse schaffen. Die Sichtbarmachung von blinden Flecken kann dazu beitragen und gezielte Entwicklungsimpulse setzen.

Insgesamt stärkt das Instrument die Anschlussfähigkeit der LevelHub-Methodik für neue Einsatzbereiche – insbesondere im Kontext von Teamentwicklung und Personalbeurteilung, wo die gezielte Beobachtung überfachlicher Kompetenzen zunehmend an Bedeutung gewinnt.

6.2 Implikationen für Teamentwicklung und Personalbeurteilung

Das entwickelte VAS-KOOP zeigt vielversprechende Ansätze für die praktische Anwendung in teambasierten Lernformaten. Im Rahmen von Teamentwicklungsworkshops bietet das Instrument einen strukturierten Zugang zur Reflexion überfachlicher Fähigkeiten. Denkbar ist etwa, dass eine spezifische Kategorie wie Teamarbeit & Kommunikation als inhaltlicher Fokus eines Tagesziels definiert wird. Die Teilnehmenden setzen sich dabei im Spiel und in der anschließenden Reflexion bewusst mit ihrem Verhalten in kollaborativen Problemlöseprozessen auseinander. Das spielerische Setting kann dabei eine gewisse Lockerheit schaffen, die es erleichtert, über eigene Verhaltensmuster zu sprechen und konstruktiv zu reflektieren.

Die Besonderheit des Instruments liegt in der Verwendung klar formulierter Verhaltensindikatoren, die als Grundlage für systematische Rückmeldungen dienen. Dies ermöglicht differenzierte Diskussionen im Anschluss an die Spielphase und erleichtert die Übertragung in reale Arbeitskontexte. Zudem lassen sich Entwicklungsverläufe über mehrere Workshops hinweg nachvollziehen und mit gezielten Massnahmen fördern.

Auch im Bereich der Personalbeurteilung eröffnet VAS-KOOP neue Möglichkeiten. Es könnte als ergänzendes Instrument in Auswahlverfahren eingesetzt werden – insbesondere zur Beobachtung überfachlicher Kompetenzen in realitätsnahen, kooperativen Settings. Voraussetzung dafür wäre, dass das eingesetzte Spiel inhaltlich zum beruflichen Anforderungsprofil passt. So könnte beispielsweise das Videospiel *Overcooked! 2* (Team17 & Ghost Town Games, 2018), das auf effektive Kommunikation und Teamkoordination angewiesen ist, in Auswahlverfahren für Berufe mit hohen Anforderungen an Kooperationsfähigkeit – etwa in der Flugsicherung oder im Operationssaal – eingesetzt werden. Zwar ist die direkte Übertragbarkeit der Ergebnisse auf den Berufsalltag begrenzt, doch können die beobachteten Verhaltensweisen als Impuls für vertiefende Gespräche und Einschätzungen dienen.

Die LevelHub-Methodik mit dem VAS-KOOP vereint spielerisches Lernen mit strukturierter Verhaltensbeobachtung und eröffnet so neue Wege für Kompetenzdiagnostik. In Teamentwicklung wie Personalbeurteilung kann es neben klassischen Methoden als innovatives Instrument beigezogen werden – vorausgesetzt, es wird kontextgerecht eingesetzt.

6.3 Perspektiven für zukünftige Forschung

Für eine Weiterentwicklung des VAS-KOOP bieten sich mehrere Ansatzpunkte an. Eine Kombination des Instruments mit weiteren diagnostischen Verfahren – etwa psychometrisch validierten Skalen zu überfachlichen Kompetenzen wie Entscheidungsfindung, Teamarbeit oder Führung – könnte die Konstruktvalidität stärken und helfen, Abweichungen zwischen Selbst- und Fremdwahrnehmung differenzierter zu verstehen. Auch eine erneute Prüfung der Reliabilität mit einer grösseren Zahl an Bewertenden erscheint sinnvoll, um die Generalisierbarkeit der bisherigen Ergebnisse zu erhöhen.

Darüber hinaus lohnt sich die Erprobung des Instruments mit weiteren kooperativen Videospiele (Farah et al., 2022). Dies könnte Aufschluss über die Skalierbarkeit des Instruments liefern – insbesondere in Hinblick auf den wachsenden Einsatzbereich als GBA.

Ein weiterer Forschungsschwerpunkt sollte auf der Erweiterung der Verhaltensindikatoren um emotionale und paraverbale Aspekte liegen – etwa sichtbare Frustration oder Tonlage. Solche Signale beeinflussen laut Mehrabian (2017) nicht nur das Verständnis der kommunikativen Inhalte, sondern prägen auch wesentlich die Beziehungsebene der Interaktion. Ihre systematische Erfassung könnte die Analyse sozialer Dynamiken im Spielkontext substanziell erweitern.

Schliesslich stellt das gemeinsame Spielen mehrerer Personen an einem Bildschirm ein bislang wenig erforschtes, aber besonders vielversprechendes Feld dar. Anders als Online-Multiplayer-Games ermöglicht dieses Setting die unmittelbare Beobachtung sozialer Verhaltensweisen. Künftige Forschung sollte daher verstärkt das Potenzial lokaler Spielsituationen nutzen, um das soziale Zusammenspiel in spielbasierten Lernszenarien besser zu verstehen und sichtbar zu machen.

6.4 Fazit

Was passiert also eigentlich, wenn Menschen gemeinsam ein Videospiele spielen – und woran erkennen wir, wie gut sie zusammenarbeiten? Mit dem VAS-KOOP liegt nun ein Instrument vor, das genau dafür einen strukturierten Rahmen bietet. Es macht kooperatives Verhalten im Spiel nicht nur beobachtbar und vergleichbar, sondern auch gezielt entwickelbar und eröffnet damit neue Wege für Diagnostik, Reflexion und Kompetenzförderung.

7 Literaturverzeichnis

- Arashi Vision Inc. (2023). Insta360 Studio (V4.6.0) [Computer Software]. Shenzhen, China: Arashi Vision Inc. Zugriff am 28.5.2025. Verfügbar unter: https://www.insta360.com/support/supportcourse?post_id=20330
- Arora, S., Miskovic, D., Hull, L., Moorthy, K., Aggarwal, R., Johannsson, H. et al. (2011). Self vs expert assessment of technical and non-technical skills in high fidelity simulation. *The American Journal of Surgery*, 202(4), 500–506. <https://doi.org/10.1016/j.amjsurg.2011.01.024>
- Bailey, C., Pearson, E., Gkatzidou, S. & Green, S. (2006). Using Video Games to Develop Social, Collaborative and Communication Skills (S. 1154–1161). Gehalten auf der EdMedia + Innovate Learning, Association for the Advancement of Computing in Education (AACE). Zugriff am 11.5.2023. Verfügbar unter: <https://www.learntechlib.org/primary/p/23149/>
- Bakeman, R. & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis (2nd Edition)*. New York, NY: Cambridge University Press. <https://doi.org/10.1017/CBO9780511527685>
- Beigi, M., Ayoobzadeh, M. & Shirmohammadi, M. (2025). Learning by gaming: nonwork-to-work enrichment among successful massive multiplayer online gamers. *Human Resource Development International*, 28(3), 446–475. Routledge. <https://doi.org/10.1080/13678868.2024.2404818>
- Bhatia, S. & Ryan, A. M. (2018). Hiring for the win: Game-based assessment in employee selection (Research in human resource management). In J.H. Dulebohn & D.L. Stone (Hrsg.), *The brave new world of eHRM 2.0*. (S. 81–110). Charlotte, NC, US: IAP Information Age Publishing. Verfügbar unter: <https://psycnet.apa.org/record/2018-32213-004>
- Bieri Buschor, C. & Schuler Braunschweig, P. (2011). Check-point Assessment Centre für angehende Lehramtsstudierende. Empirische Befunde zur prognostischen Validität und zur Übereinstimmung von Selbst- und Fremdeinschätzung eignungsrelevanter Merkmale. *Zeitschrift für Pädagogik*, 57(5), 695–710. Weinheim: Beltz. <https://doi.org/10.25656/01:8773>
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation* (Springer-Lehrbuch). Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-540-33306-7>
- Braun, E. M. P., Sheikh, H. & Hannover, B. (2011). Self-rated competences and future vocational success: A longitudinal study. *Assessment & Evaluation in Higher Education*, 36(4), 417–427. United Kingdom: Taylor & Francis. <https://doi.org/10.1080/02602938.2010.534762>
- Brimble, M. (2008). Skills assessment using video analysis in a simulated environment: an evaluation. *Paediatric Nursing*, 20(7), 26–31. <https://doi.org/10.7748/paed2008.09.20.7.26.c6705>

- Campbell, L. M., Howie, J. G. & Murray, T. S. (1995). Use of videotaped consultations in summative assessment of trainees in general practice. *The British Journal of General Practice: The Journal of the Royal College of General Practitioners*, 45(392), 137–141. Verfügbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1239174/>
- Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A. & Hogan, R. (2016). New Talent Signals: Shiny New Objects or a Brave New World? *Industrial and Organizational Psychology*, 9(3), 621–640. <https://doi.org/10.1017/iop.2016.6>
- Clausen, M., Möller, J. & Strauß, B. (2000). Einflüsse auf das Ankreuzverhalten bei der Self-Monitoring-Skala. *Situationsspezifität in pädagogischen Handlungsfeldern* (S. 81–100). Studien-Verlag. Zugriff am 28.5.2025. Verfügbar unter: https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2102686
- da Conceição, V. F. P., Mendes, J. B., Teodoro, M. F. & Dahlman, J. (2019). Validation of a Behavioral Marker System for Rating Cadet's Non-Technical Skills. *TransNav, International Journal on Marine Navigation and Safety of Sea Transportation*, 13(1), 89–96. <https://doi.org/10.12716/1001.13.01.08>
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T. & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), 661–686. <https://doi.org/10.1016/j.compedu.2012.03.004>
- Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York, NY: Harper & Row.
- Dewe, B. (2009). Reflexive Sozialarbeit im Spannungsfeld von evidenzbasierter Praxis und demokratischer Rationalität – Plädoyer für die handlungslogische Entfaltung reflexiver Professionalität. In R. Becker-Lenz, S. Busse, G. Ehlert & S. Müller (Hrsg.), *Professionalität in der Sozialen Arbeit: Standpunkte, Kontroversen, Perspektiven* (S. 89–109). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91512-8_5
- Dietz, A. S., Pronovost, P. J., Benson, K. N., Mendez-Tellez, P. A., Dwyer, C., Wyskiel, R. et al. (2014). A systematic review of behavioural marker systems in healthcare: what do we know about their attributes, validity and application? *BMJ Quality & Safety*, 23(12), 1031–1039. <https://doi.org/10.1136/bmjqs-2013-002457>
- Ellison, L. J., Johnson, T. M., Tomczak, D., Siemsen, A. & Gonzalez, M. F. (2020). Game on! Exploring reactions to game-based selection assessments. *Journal of Managerial Psychology*, 35(4), 241–254. Emerald Publishing Limited. <https://doi.org/10.1108/JMP-09-2018-0414>
- Falchikov, N. & Boud, D. (1989). Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Research*, 59(4), 395–430. <https://doi.org/http://dx.doi.org/10.3102/00346543059004395>
- Farah, Y. A., Dorneich, M. C. & Gilbert, S. B. (2022). Evaluating Team Metrics in Cooperative Video Games. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 66(1), 70–74. <https://doi.org/10.1177/1071181322661240>

- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327–358. US: American Psychological Association. <https://doi.org/10.1037/h0061470>
- Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N. & Patey, R. (2003). Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system†. *BJA: British Journal of Anaesthesia*, 90(5), 580–588. <https://doi.org/10.1093/bja/aeg112>
- Flin, R. & Martin, L. (2001). Behavioral markers for crew resource management: A review of current practice. *The International Journal of Aviation Psychology*, 11(1), 95–118. US: Lawrence Erlbaum. https://doi.org/10.1207/S15327108IJAP1101_6
- Galuske, M. & Müller, C. W. (2002). Handlungsformen in der Sozialen Arbeit. In W. Thole (Hrsg.), *Grundriss Soziale Arbeit: Ein einführendes Handbuch* (S. 485–508). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-91357-9_27
- Goffman, E. (1959). *The presentation of self in everyday life*. Oxford, England: Doubleday.
- Gomez, M. J., Ruipérez-Valiente, J. A. & Clemente, F. J. G. (2023). A Systematic Literature Review of Game-Based Assessment Studies: Trends and Challenges. *IEEE Transactions on Learning Technologies*, 16(4), 500–515. <https://doi.org/10.1109/TLT.2022.3226661>
- Granic, I., Lobel, A. & Engels, R. C. M. E. (2014). The benefits of playing video games. *American Psychologist*, 69(1), 66–78. <https://doi.org/10.1037/a0034857>
- Griffin, P. & Care, E. (2015). *Assessment and Teaching of 21st Century Skills: Methods and Approach*. Dordrecht: Springer Netherlands. <https://doi.org/10.1007/978-94-017-9395-7>
- Haidet, K. K., Tate, J., Divirgilio-Thomas, D., Kolanowski, A. & Happ, M. B. (2009). Methods to Improve Reliability of Video Recorded Behavioral Data. *Research in nursing & health*, 32(4), 465–474. <https://doi.org/10.1002/nur.20334>
- Hays, R. B., Jones, B. F., Adkins, P. B. & McKain, P. J. (1990). Analysis of videotaped consultations to certify competence. *The Medical Journal of Australia*, 152(11), 609–611. <https://doi.org/10.5694/j.1326-5377.1990.tb125395.x>
- Hennessy, J., Mabey, B. & Warr, P. (1998). Assessment centre observation procedures: An experimental comparison of traditional, checklist and coding methods. *International Journal of Selection and Assessment*, 6(4), 222–231. United Kingdom: Blackwell Publishing. <https://doi.org/10.1111/1468-2389.00093>
- Higgins, E. T. & Bargh, J. A. (1987). Social cognition and social perception. *Annual Review of Psychology*, 38, 369–425. US: Annual Reviews. <https://doi.org/10.1146/annurev.ps.38.020187.002101>
- Höft, S. & Kersting, M. (2018). Anforderungsprofil, Verhaltensbeobachtung und Verhaltensbeurteilung. In Diagnostik- Und Testkuratorium (Hrsg.), *Personalauswahl kompetent gestalten* (S. 27–63). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-53772-5_2
- Husein, M. H., Ow, S. H., Elaish, M. M. & Jensen, E. O. (2022). Digital game-based learning in K-12 mathematics education: a systematic literature review. *Education and Information Technologies*, 27(2), 2859–2891. <https://doi.org/10.1007/s10639-021-10721-x>

- Hussy, W., Schreier, M. & Echterhoff, G. (2013). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (Springer-Lehrbuch) (2. Aufl. 2013.). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-34362-9>
- James, L. R., Demaree, R. G. & Wolf, G. (1993). rwg: An assessment of within-group interrater agreement. *Journal of Applied Psychology*, 78(2), 306–309. US: American Psychological Association. <https://doi.org/10.1037/0021-9010.78.2.306>
- Kanning, U. P. (2019). *Diagnostik für Führungspositionen*. Hogrefe. <https://doi.org/10.1026/02946-000>
- Kleinmann, M. (2016). *Assessment-Center*. Göttingen: Hogrefe Verlag.
- Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. US: American Psychological Association. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Landers, R. N. (2015). An Introduction to Game-based Assessment: Frameworks for the Measurement of Knowledge, Skills, Abilities and Other Human Characteristics using Behaviors Observed within Videogames. *International Journal of Gaming and Computer-Mediated Simulations*, 7(4).
- Landers, R. N. & Sanchez, D. R. (2022). Game-based, gamified, and gamefully designed assessments for employee selection: Definitions, distinctions, design, and validation. *International Journal of Selection and Assessment*, 30(1), 1–13. <https://doi.org/10.1111/ijsa.12376>
- Luft, J. & Ingham, H. (1955). The Johari window, a graphic model of interpersonal awareness. *Proceedings of the western training laboratory in group development*. Los Angeles: University of California, Los Angeles. Zugriff am 28.5.2025. Verfügbar unter: https://ombuds.columbia.edu/sites/default/files/content/pics/30%20Anniv/The%20Johari%20window_A%20graphic%20model%20of%20awareness%20in%20interpersonal%20relations.pdf
- Manser, T., Schmutz, J. & Perry, J. (2018). *Verhalten ist messbar: Behavioural Marker Systeme und Kompetenzentwicklung*. Springer. Verfügbar unter: <http://hdl.handle.net/11654/26766>
- Mehrabian, A. (2017). *Nonverbal Communication*. New York: Routledge. <https://doi.org/10.4324/9781351308724>
- Mitchell, L., Flin, R., Yule, S., Mitchell, J., Coutts, K. & Youngson, G. (2012). Evaluation of the Scrub Practitioners' List of Intraoperative Non-Technical Skills system. *International Journal of Nursing Studies*, 49(2), 201–211. <https://doi.org/10.1016/j.ijnurstu.2011.08.012>
- Moosbrugger, H. & Kelava, A. (2020). Qualitätsanforderungen an Tests und Fragebogen („Gütekriterien“). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 13–38). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-61532-4_2

- Moreno, R. & Mayer, R. (2007). Interactive Multimodal Learning Environments. *Educational Psychology Review*, 19(3), 309–326. <https://doi.org/10.1007/s10648-007-9047-2>
- Obermann, C. (2002). *Assessment Center*. Wiesbaden: Gabler Verlag.
<https://doi.org/10.1007/978-3-322-92038-6>
- OECD. (2018). Future of Education and Skills 2030/2040. *OECD*. Zugriff am 9.4.2025. Verfügbar unter: <https://www.oecd.org/en/about/projects/future-of-education-and-skills-2030.html>
- OECD. (2023). OECD Skills Outlook 2023. *OECD*. Zugriff am 29.4.2025. Verfügbar unter: https://www.oecd.org/en/publications/oecd-skills-outlook-2023_27452f29-en.html
- Ohlms, M. L., Melchers, K. G. & Kanning, U. P. (2024). Playful personnel selection: The use of traditional versus game-related personnel selection methods and their perception from the recruiters' and applicants' perspectives. *International Journal of Selection and Assessment*, 32(3), 381–398. <https://doi.org/10.1111/ijsa.12466>
- Ouellette, J. A. & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin*, 124(1), 54–74. <https://doi.org/10.1037/0033-2909.124.1.54>
- Podsakoff, P. M., MacKenzie, S. B., Moorman, R. H. & Fetter, R. (1990). Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *The Leadership Quarterly*, 1(2), 107–142. Netherlands: Elsevier Science. [https://doi.org/10.1016/1048-9843\(90\)90009-7](https://doi.org/10.1016/1048-9843(90)90009-7)
- Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L. & Lazzara, E. H. (2015). Understanding and Improving Teamwork in Organizations: A Scientifically Based Practical Guide. *Human Resource Management*, 54(4), 599–622. <https://doi.org/10.1002/hrm.21628>
- Schaubroeck, J., Lam, S. S. K. & Cha, S. E. (2007). Embracing transformational leadership: Team values and the impact of leader behavior on team performance. *Journal of Applied Psychology*, 92(4), 1020–1030. <https://doi.org/10.1037/0021-9010.92.4.1020>
- Schmidt-Atzert, L., Krumm, S. & Amelang, M. (2021). *Psychologische Diagnostik*. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-662-61643-7>
- Schulz, B. (2008). The importance of soft skills: Education beyond academic knowledge. *Journal of Language and Communication*, 2, 146–154. Verfügbar unter: <https://ir.nust.na/server/api/core/bitstreams/53056486-f186-4d65-96b4-70fe0b9f8e91/content>
- Smith, P. C. & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 47(2), 149–155. US: American Psychological Association.
<https://doi.org/10.1037/h0047060>
- Sweetser, P. & Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment*, 3(3), 3–3. <https://doi.org/10.1145/1077246.1077253>
- Team17 & Ghost Town Games. (2018). *Overcooked! 2* [Videospiele]. Wakefield, England. Verfügbar unter: <https://www.team17.com/games/overcooked-2>
- The jamovi project. (2025). Jamovi (Version 2.4) [Computer Software]. Zugriff am 28.5.2025. Verfügbar unter: <https://www.jamovi.org/>

- Tinsley, H. E. & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, 22(4), 358–376. <https://doi.org/10.1037/h0076640>
- Trespalacios, J., Chamberlin, B. & Gallagher, R. R. (2011). Collaboration, Engagement & Fun: How Youth Preferences in Video Gaming can Inform 21st Century Education. *TechTrends*, 55(6), 49–54. <https://doi.org/10.1007/s11528-011-0541-5>
- Undre, S., Sevdalis, N., Healey, A. N., Darzi, A. & Vincent, C. A. (2007). Observational Teamwork Assessment for Surgery (OTAS): Refinement and Application in Urological Surgery. *World Journal of Surgery*, 31(7), 1373–1381. <https://doi.org/10.1007/s00268-007-9053-z>
- Wernimont, P. F. & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, 52(5), 372–376. US: American Psychological Association. <https://doi.org/10.1037/h0026244>
- World Economic Forum. (2025). The Future of Jobs Report 2025. *World Economic Forum*. Zugriff am 13.4.2025. Verfügbar unter: <https://www.weforum.org/publications/the-future-of-jobs-report-2025/in-full/3-skills-outlook/>
- Yule, S., Flin, R., Paterson-Brown, S. & Maran, N. (2006). Non-technical skills for surgeons in the operating room: a review of the literature. *Surgery*, 139(2), 140–149. Elsevier. <https://doi.org/10.1016/j.surg.2005.06.017>
- Yule, S., Flin, R., Paterson-Brown, S., Maran, N. & Rowley, D. (2006). Development of a rating system for surgeons' non-technical skills. *Medical Education*, 40(11), 1098–1104. <https://doi.org/10.1111/j.1365-2929.2006.02610.x>
- Yule, S., Gupta, A., Gazarian, D., Geraghty, A., Smink, D. S., Beard, J. et al. (2018). Construct and criterion validity testing of the Non-Technical Skills for Surgeons (NOTSS) behaviour assessment tool using videos of simulated operations. *The British Journal of Surgery*, 105(6), 719–727. <https://doi.org/10.1002/bjs.10779>

8 Abbildungsverzeichnis

Abbildung 1. Ablauf eines typischen LevelHub-Tages: Vom definierten Ziel über die Auswahl passender Spiele, Spielphasen und Reflexionen bis hin zum individuellen Take-away (LevelHub).....	8
Abbildung 2. Links: Modellhafte Darstellung einer klassischen Game-Session, wobei mehrere Teams gleichzeitig spielen und Beobachtungen gemacht werden. Rechts: eine symbolische Darstellung eines erfolgreichen kooperativen Prozesses (LevelHub).	9
Abbildung 3. Vierstufiges Wirkmodell der Vorgehensweise – von der Beobachtung über Reflexion und Anwendung bis zum Transfer in den Alltag (LevelHub).	9
Abbildung 4. Originale Skill-Taxonomie des NOTSS (Yule, Flin, Paterson-Brown, Maran & Rowley, 2006).	13
Abbildung 5. Verhaltensindikatoren für den Kontext eines kooperativen Videospiele, die mittels CIT gesammelt und den negativen oder positiven Polen zugewiesen wurden (eigene Abbildung).....	18
Abbildung 6. Ausschnitt aus dem VAS-KOOP zur Selbsteinschätzung mit angepasster Bewertungsskala (eigene Abbildung).	19
Abbildung 7. Ausschnitt aus dem VAS-KOOP zur Fremdeinschätzung mit Kategorien, Definitionen sowie positiven und negativen Verhaltensindikatoren pro Dimensionen (eigene Abbildung).	19
Abbildung 8. Ablauf der Datenerhebung mit den zentralen und videodokumentierten Phasen (eigene Abbildung).	21
Abbildung 9. Leveldesign des durchgeführten Level 1 in Overcooked! 2 (Team17 & Ghost Town Games, 2018) mit Darstellung sämtlicher Utensilien, Bestellungen, Punktestand und Zeitangaben (Screenshot).	22
Abbildung 10. Station der Datenerhebung mit eingerichteter Konsole, Videokamera und angeschriebenen Sitzplätzen (eigenes Foto).....	23
Abbildung 11. Mittelwerte der gemessenen Kategorien: Entscheidungsfindung, Kommunikation & Teamarbeit sowie Führung über die drei Perspektiven hinweg; N = 30.	27
Abbildung 12. Mittelwerte der Bewertungen A und B getrennt nach den Kategorien Entscheidungsfindung, Kommunikation & Teamarbeit und Führung; N = 30 (erstellt mit datab.de).....	30

9 Tabellenverzeichnis

Tabelle 1 Adaptierte Taxonomie mit den drei Hauptkategorien und jeweils drei Unterdimensionen	16
Tabelle 2 Zentrale Evaluationskriterien der Untersuchung mit Bewertungsfragen und Datenquellen	25
Tabelle 3 Darstellung der Mittelwerte und Standardabweichungen sämtlicher gemessenen Dimensionen und Kategorien über die drei Perspektiven hinweg	28
Tabelle 4 Reliabilität des VAS-KOOP-Instruments zur Bewertung überfachlicher Fähigkeiten in drei Kategorien	29
Tabelle 5 Übereinstimmung der erhobenen Kategorien innerhalb der Gruppe (r_{wg}) über drei Vergleichspaare hinweg.....	30

10 Hilfsmittelverzeichnis mit Verwendungszweck

Verzeichnis der verwendeten Hilfsmittel

Assistenzsystem	Teile / Stelle(n) in der Arbeit	Einsatz
DeepL	Divers	Übersetzung von Studien
ChatGPT	Recherche	Sammeln und Prüfen von ersten Studien und Literaturvorschlägen
ChatGPT	Divers	Übersetzungen von Abstracts, Inspiration für Satz-Formulierungen, Überprüfung auf Rechtschreibfehler