



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Bericht zur Pilotevaluation

Tutoring for all - Lesen mit dem Turbo-Team



Thorsten Klinger & Ingrid Gogolin

Universität Hamburg

Fakultät für Erziehungswissenschaft

Arbeitsgruppe Interkulturelle und International Vergleichende Erziehungswissenschaft

Forschungsschwerpunkt „Literacy in Diversity Settings – LiDS“

Von-Melle-Park 8

20146 Hamburg

26. August 2024

Bitte zitieren Sie diesen Text wie folgt:

Klinger, Thorsten; Gogolin, Ingrid (2024): Tutoring for all – Lesen mit dem Turbo-Team.

Bericht zur Pilotevaluation. Hamburg

(Universität Hamburg; <https://www.tutoringforall.uni-hamburg.de/>)

Inhalt

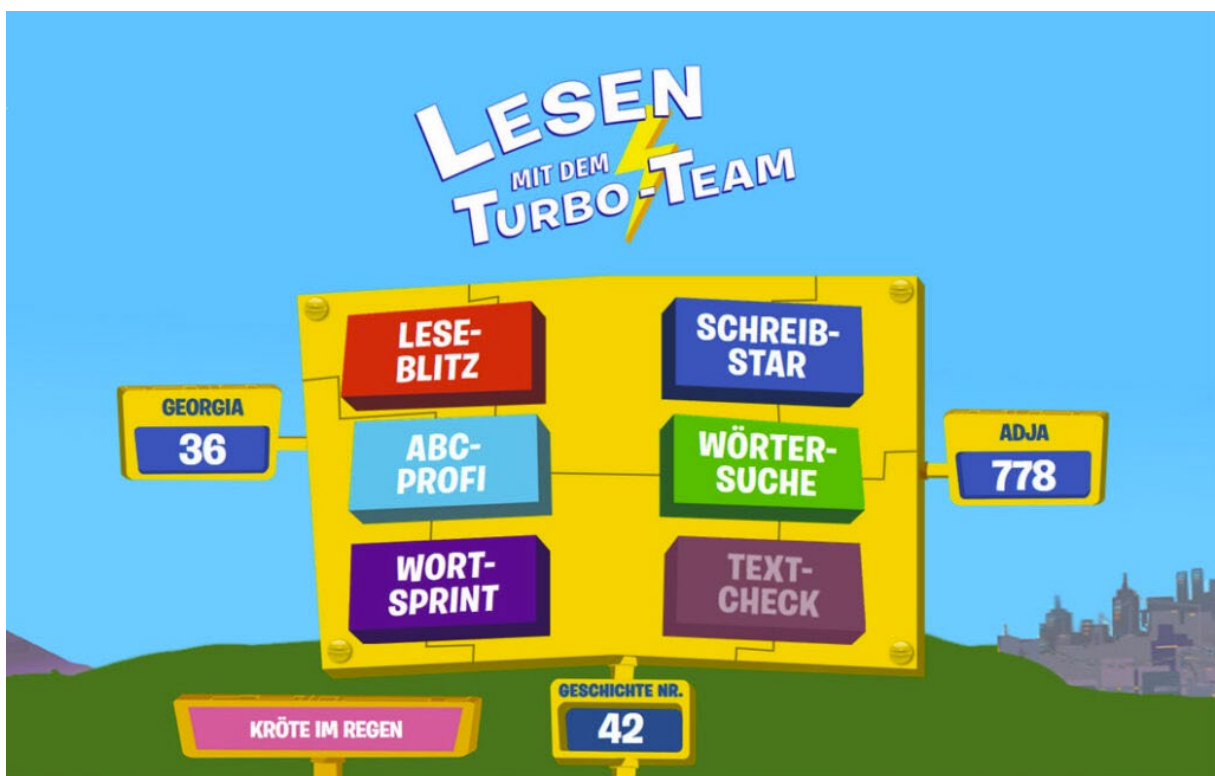
Vorbemerkung.....	5
1. Einleitung.....	7
2. Das Konzept der Pilotevaluation.....	8
2.1. Design der Evaluation.....	8
2.1.1. Wirksamkeitsprüfung.....	8
2.1.2. Machbarkeit und Akzeptanz.....	10
3. Durchführung der tentativen Wirksamkeitsprüfung.....	11
4. Analysestrategie und deskriptive Befunde.....	14
4.1. Getestete Lesefähigkeiten.....	14
4.2. Resultate der Tests, Zusammensetzung der Testgruppen und Folgen für die Regressionsmodelle.....	15
5. Die Resultate im fairen Vergleich.....	18
5.1. Ergebnisse.....	19
5.2. Diskussion.....	21
6. Durchführung und Resultate der Tutor:innen-Befragung.....	23
6.1. Teilnehmende an der Befragung.....	24
6.2. Einschätzungen und Erfahrungen der Tutor:innen.....	26
6.2.1. Einstellung zu TFA und Einschätzung der Wirksamkeit.....	27
6.2.2. Praxiserfahrungen, Durchführbarkeit und wahrgenommene Eignung von TFA...30	
6.3. Diskussion der Befragungsergebnisse.....	38
Literatur.....	40
Anhang I: Power-Analyse mit G*Power.....	43
Anhang II: Antworten auf offene Fragen in der Tutor:innenbefragung.....	44

Thorsten Klinger & Ingrid Gogolin
Universität Hamburg

Bericht zur Pilotevaluation: Tutoring for all – Lesen mit dem Turbo-Team

Vorbemerkung

Das Projekt „Tutoring for all – Lesen mit dem Turbo-Team“ (<https://tutoringforall.de/>) ist Teil einer Zusammenarbeit zwischen der „Tutoring for all gUG“ (Ekkehard Thümler) und der Arbeitsgruppe Interkulturelle und International Vergleichende Erziehungswissenschaft der Universität Hamburg (Gogolin, Klinger und andere, <https://www.ew.uni-hamburg.de/ueber-die-fakultaet/personen/gogolin.html>), die zum Forschungsschwerpunkt Literacy in Diversity Settings (LiDS) gehört. Gemeinsam haben die beiden Partner eine finanzielle Unterstützung dafür erhalten, einen Prototyp für die Förderung der Literalität von Grundschüler(inne)n in Deutschland zu entwickeln. Diese finanzielle Unterstützung erfolgte im Rahmen des „Innovationsprogramms für Geschäftsmodelle und Pionierlösungen (IGP)“ des Bundesministeriums für Wirtschaft und Energie. Gegenstand der Unterstützung für die Arbeitsgruppe der Universität Hamburg war die prozessbegleitende formative Evaluation des entstehenden Produkts: der Plattform „Lesen mit dem Turbo Team“:



Dem Bundesministerium für Wirtschaft und Energie sowie dem Team des VDI/VDE Innovation + Technik GmbH, das die operative Umsetzung der Projektförderung konstruktiv begleitete, danken wir für die finanzielle Unterstützung und die sehr gute Zusammenarbeit im Prozess. Unser ganz besonderer Dank gilt Ekkehard Thümmler, dem Initiator von „Lesen mit dem Turbo-Team“, der uns auf die interessante Reise der Entwicklung mitgenommen hat; den Partnerinnen von InSL e.v., allen voran Kristina Edel und Sümeyye Balci; dem Tutorinnenteam, den Schulen und den großartigen Kindern, die sich mit uns auf das Wagnis der Pilotevaluation eingelassen haben.

1. Einleitung

Ziel des Vorhabens war und ist es, auf der Grundlage des US-amerikanischen Programms "Tutoring with the Lightning Squad" in Zusammenarbeit mit der Tutoring for All gUG eine passgenaue Etablierung des Programms "Tutoring for All (TFA) – Lesen mit dem Turbo-Team" zur Förderung der Literalität von Grundschüler(inne)n in Deutschland anzustreben. Dieses Programm soll zum Ausgleich von zumeist herkunftsbedingten Nachteilen beim Lesen- (und Schreiben-)lernen beitragen. Aufgabe des Teams der Universität Hamburg war die Realisierung einer begleitenden Pilotevaluation zu Fragen von Usability und Nutzerakzeptanz. Die Evaluation sollte zugleich Informationen über die Förderqualität des Programms erkennen lassen, die in die Weiterentwicklung eingespeist werden können.

Die Untersuchung der Universität Hamburg wurde als Pilotevaluation konzipiert und durchgeführt, weil sie im Prozess der Entwicklung bereits Hinweise auf potenzielle Wirkungen des Tutorings erbringen sollte. Die angestrebten Anhaltspunkte auf die Qualität und Wirksamkeit des Vorgehens sollten vor allem mögliche Nachsteuerungsbedarfe erkennen lassen. Die Untersuchung war also bewusst so angelegt, dass dem Entwicklungsteam des Projekts bereits zu einem frühen Zeitpunkt Rückmeldungen dazu zurückgespiegelt werden können, ob der gewählte Weg in die richtige Richtung führt. Auch wenn in Hinblick auf die Einschätzung der Wirksamkeit der Förderung Elemente einer summativen Evaluation (Testverfahren, Kontrollgruppendesign) eingeflossen sind, sind Anspruch und Schwerpunkt der Pilotevaluation eindeutig im Bereich einer den Entwicklungsprozess begleitenden, also formativen Evaluation zu verorten.

Im Vergleich mit einer summativen Evaluation der Maßnahme im Sinne einer regulären Wirkungsevaluation sind die hier berichteten Untersuchungsergebnisse dementsprechend nur eingeschränkt aussagekräftig. Sie stehen insbesondere unter dem Vorbehalt einer kleinen Fallzahl in der Untersuchung und der Tatsache, dass keine randomisierte Zuweisung der Schülerinnen und Schüler zur Interventions- und zur Kontrollgruppe stattfinden konnte.

Gegenstand des vorliegenden Berichts über die Pilotevaluation ist die Darstellung des Evaluationskonzepts und der Ergebnisse, die zur formativen Bewertung der Wirksamkeit des Tutorings sowie zur Bewertung der Machbarkeit und zur Akzeptanz durch die beteiligten Tutor:innen erzielt wurden.

Der Bericht bezieht sich auf die Versionen des TFA-Programms, die zur Zeit der Tests und der Befragung eingesetzt wurden. Erweiterungen und Änderungen, die inzwischen stattgefunden haben, konnten nicht berücksichtigt werden. Da das TFA-Entwicklungsteam regelmäßig und zeitnah über die Ergebnisse der Pilotevaluation auf dem Laufenden gehalten wurde, konnten die Resultate in die Weiterentwicklung einfließen.

2. Das Konzept der Pilotevaluation

Das prinzipielle Ziel einer wissenschaftlichen Evaluation im Bereich der Erprobung von Förder- oder Bildungskonzepten besteht darin, die Wirksamkeit des Konzepts zu bewerten, um sicherzustellen, dass die beabsichtigten Ergebnisse erzielt werden und den Bedürfnissen der jeweiligen Zielgruppen entsprechen. Als wissenschaftliche Evaluation stützt sich die jeweilige Beurteilung auf Daten und Erkenntnisse, die auf der systematischen Anwendung wissenschaftlicher Methoden beruhen (Bortz & Döring 2002, 102; Rossi et al. 1988, 3). Dabei ist zu unterscheiden zwischen einer auf die Projektergebnisse hin ausgerichteten, nachgelagerten summativen Evaluation und einer für die Projektsteuerung nutzbaren, prozessbegleitenden formativen Evaluation (Haack & Haß 2024). Das frühe Entwicklungsstadium des Projekts Tutoring for All (TFA) verlangte eindeutig nach einer Prozessbegleitung. Diese sollte aber gleichzeitig eine Vorschau auf die potentielle Wirksamkeit des Programms im summativen Sinne ermöglichen, um zu einem frühen Zeitpunkt gegebenenfalls Anhaltspunkte für die Nachsteuerung dieses Kernbereichs des Projekts bereitstellen zu können.

Als Resultat dieser Überlegungen entschied sich das Projektteam der Universität Hamburg in Absprache mit der Leitung von TFA für eine Pilotevaluation mit formativem Charakter und summativen Elementen, die aber noch keine abschließende Bewertung der Wirksamkeit erlauben oder anstreben, sondern Anhaltspunkte dafür bieten, inwieweit die deutsche Version des Programms sich in die Richtung entwickelt, die durch die positiven Evaluationsergebnisse des US-amerikanischen „Originals“ vorgezeichnet ist (Chambers et al., 2008; Madden & Slavin 2017). Einig waren sich die Projektbeteiligten darüber, dass die Universität Hamburg als externe Evaluation agiert, um die Unabhängigkeit der Evaluator:innen und der Resultate zu gewährleisten. Die Rückspiegelung der Ergebnisse und deren Diskussion sollten auch als Basis für weitere interne Evaluationsaktivitäten des TFA-Teams genutzt werden können.

2.1. Design der Evaluation

Die Pilotevaluation besteht aus zwei Elementen:

- (1) der Bewertung des *Nutzens*, festgemacht an der Wirksamkeit der Förderung der Lesefähigkeiten bei den getesteten Grundschulkindern;
- (2) der Bewertung der *Nützlichkeit*, festgemacht an Informationen über die Machbarkeit und Akzeptanz des Einsatzes der Förderplattform im institutionellen Umfeld.

Beide Teile der Evaluation sind aufeinander bezogen, da die finale Bewertung der Nützlichkeit im Sinne einer Kosten-Nutzen-Relation maßgeblich mit Blick auf die erzielten Wirkungen des Tutorings erfolgt.

2.1.1. Wirksamkeitsprüfung

Zu den Zielen der Hinführung zum Lesen in der Grundschule gehört zunächst, dass Kinder flüssig mit Texten umgehen können. Darauf aufbauend, geht es zunehmend

darum, sie zum „Lesen, um zu lernen“ zu führen. Schnelles und flüssiges Lesen bilden das Handwerkszeug dafür, aus Texten verschiedener Genres und zunehmender Komplexität zügig Informationen zu entnehmen, diese zu interpretieren und zu bewerten (Rosebrock & Nix 2020, 26ff). Die Grundlagen dieser Fähigkeiten sollten etwa bis zur dritten Klassenstufe gelegt sein. Das Evaluationsdesign knüpfte an diesem Stand der Erkenntnisse zur Aneignung von Lesekompetenz an und sah die Wirksamkeitsprüfung an Schüler:innen der zweiten und dritten Klassenstufe vor.

Zur Bestimmung der Wirksamkeit des Tutorings sollten die teilnehmenden Kinder mit anerkannten Verfahren dahingehend getestet werden, ob nach der Teilnahme am Programm eine Verbesserung ihrer Lesefähigkeit feststellbar ist, die sich signifikant von der vergleichbarer Kinder ohne Tutoring unterscheidet. Die Testverfahren sollten objektive, reliable und valide Messungen von grundlegenden altersgemäßen Leseleistungen garantieren und die Leistungen unabhängig, d.h. ohne Rückgriff auf die im Tutoring-Programm implementierten Analyse-Tools erfassen.

Als Mindestanforderung für die Feststellung von Fortschritten in der Lesefähigkeit gilt ein Prä-Post-Kontrollgruppen-Design, das eine Testung von Kindern vor Beginn und nach Abschluss einer hinreichend langen Phase einer Intervention sowie die parallele zweimalige Testung von Kindern vorsieht, die nicht an der Intervention teilnehmen.

Zur Bestimmung längerfristiger Effekte wäre ein Prä-Post-Follow-Up-Kontrollgruppen-design wünschenswert. Zugleich aber konkurrierten im vorliegenden Fall das Interesse am formativen Charakter der Pilotevaluation – also der Möglichkeit einer raschen Rückmeldung an die Programmentwicklung – und das Interesse an Informationen zur Nachhaltigkeit der Effekte. Aufgrund dieser Rahmenbedingungen der Entwicklung des Tutoring-Programms wurde für die Pilotevaluation das sparsame Prä-Post-Kontrollgruppen-Verfahren gewählt.

Das zunächst ins Auge gefasste Evaluationsdesign sah vor, die Kinder spätestens nach der Ersttestung nach dem Zufallsprinzip einer Experimental- und einer Kontrollgruppe zuzuweisen, wobei die Experimentalgruppe in der ersten Phase am Tutoring teilnehmen und die Kontrollgruppe das Tutoring nach der zweiten Testung nachholen sollte. Zum Zeitpunkt der Testungen konnte aufgrund der Teilnehmendenzahl und der begrenzten regionalen Verbreitung von TFA keine echte Zufallszuweisung erfolgen. Vielmehr musste die Kontrollgruppe durch Kinder einer Schule ergänzt werden, an der mangels erforderlicher technischer Ausstattung kein Tutoring mit der elektronischen Plattform durchgeführt werden konnte.

Bereits in der Planungsphase des Projekts war deutlich geworden, dass sich die Testresultate aufgrund der zu erwartenden Fallzahlen von 60 bis 100 Kindern nicht hinreichend statistisch absichern lassen. Bei Annahme einer mittleren Effektstärke ist nach Poweranalyse (Faul et al. 2007) eine Fallzahl von insgesamt mindestens 269 Personen erforderlich, damit ein signifikanter Effekt mit einer Wahrscheinlichkeit von 80% nachgewiesen werden kann (vgl. Anhang I, S. 43). Die Planung des

Evaluationsdesigns im Verständnis eines formativen Assessments erlaubt jedoch, Anhaltspunkte für die Wirksamkeit der Intervention zu ermitteln, und damit zugleich Hinweise auf mögliche Nachsteuerungsbedarfe. Zudem konnte davon ausgegangen werden, dass die Stichprobe durch spätere sukzessive Erhebungen kumulativ vergrößert werden kann. Es könnte also eine weitere Evaluation angeschlossen werden, wenn sich abzeichnet, dass vergleichbare positive Effekte mit der Maßnahme erzielt werden, wie sie in den vorliegenden US-amerikanischen Evaluationen ermittelt wurden (Madden & Slavin, 2017; Ross et al. 2017). In diesem Fall sollte auch eine Nachhaltigkeitsprüfung mithilfe einer Follow-Up-Erhebung erwogen werden. Solche Erweiterungen waren im Rahmen des laufenden Evaluationsvorhabens allerdings mangels personeller, zeitlicher und finanzieller Ressourcen nicht möglich.

2.1.2. Machbarkeit und Akzeptanz

Für die Bewertung der Machbarkeit und Akzeptanz des Einsatzes des Tutorings im institutionellen Umfeld war zunächst eine Befragung der Tutor:innen vorgesehen, die sich auf die folgenden Themen erstrecken sollte:

Mit Bezug auf die Teilnehmerinnen und Teilnehmer am Tutoring sollte eine Einschätzung der Resonanz der Schüler:innen und der Passung der Aufgabenstellungen für die Zielgruppe vorgenommen werden. Berichtet werden sollte zudem über wahrgenommene Über- oder Unterforderung der Teilnehmenden und das Anhalten der Teilnahmebereitschaft über die Zeit.

Mit Bezug auf die eigene Wahrnehmung und Beurteilung des Tutorings durch die Tutor:innen sollte eine Einschätzung zur Frage vorgenommen werden, ob das vorbereitende Training zu hinreichenden Kenntnissen über Hintergrund und Durchführung der Maßnahme geführt habe. Zudem sollten Erfahrungen hinsichtlich der Durchführbarkeit der Aufgaben und auftretender Problemfälle bei der Durchführung berichtet werden. Weitere Fragen richteten sich auf die Inanspruchnahme externer Unterstützung, das Auftreten von Konfliktpunkten und auf Optimierungswünsche auf Seiten der Tutorinnen oder Tutoren.

Vorgesehen waren zudem Befragungen von Lehrkräften und Schulleitungen nach dem wahrgenommenen Nutzen von TFA. Ferner sollten Einschätzungen des eigenen Aufwands und der Kosten in Relation zum Nutzen erfolgen. Die Realisierung dieser Befragung hätte erfordert, dass hinreichend Erfahrung mit der Maßnahme gewonnen werden konnte. Die Implementation des Tutorings an Schulen nahm jedoch erst gegen Ende des Evaluationsprojekts an Fahrt auf. Da in dem verbleibenden knappen Zeitraum keine hinreichenden Erfahrungen seitens der institutionellen Teilnehmenden erwartet werden konnten, wurde auf die diesbezüglichen Befragungen verzichtet. Das Instrumentarium für die Befragungen wurde jedoch entwickelt.

3. Durchführung der tentativen Wirksamkeitsprüfung

Nachdem das Programm *Lesen mit dem Turbo-Team* einsatzbereit war und ein Team von Tutor:innen geschult zur Verfügung stand, wurde die Durchführung der Pilotevaluation so eingerichtet, dass die Prä- und Posterhebungen vor den Sommerferien abgeschlossen werden konnten. Zu diesem Zeitpunkt waren die Implementation und die Schulung von Tutor:innen auf die Fördereinrichtung InSL e.V. beschränkt, einen gemeinnützigen Verein zur Förderung von Sprachbildung in hessischen Gemeinden (<https://www.sprachelesenlernen.de/>). Mehrere Schulen waren durch die Zuweisung von Schüler:innen mit Bedarf an Leseförderung an die Angebote von InSL e.V. beteiligt. Dadurch kamen für die Kontrollgruppe nur von den Schulen zugewiesene Schüler:innen in Frage, die an einer herkömmlichen Leseförderung des Vereins teilnahmen. Schüler:innen mit Förderbedarf ohne eine Förderung standen also nicht zur Verfügung. Die Herausforderung für die Bewährung des TFA-Programms war somit größer als die, die bei einer üblichen Intervention der Fall wäre: Die Wirksamkeit des Programms wurde gegen eine alternative Förderung getestet, die methodisch konventionell, d.h. ohne systematische Unterstützung durch die elektronische Plattform realisiert wird.

Für die Teilnahme an den Tests standen Schüler:innen der Klassenstufen 2 und 3 aus sechs Schulen zur Verfügung, die von ihren Lehrerinnen wegen Förderbedarfs im Lesen an den Förderverein InSL e.V. verwiesen wurden. Alle Teilnehmenden hatten neben dem Deutschen eine andere Herkunftssprache. Welche dies war, wurde aus Gründen des Datenschutzes nicht erhoben. Für die Teilnahme an den Tests war das Einverständnis der Eltern einzuholen, die zuvor über die Tests und das Datenschutzmodell der Evaluation informiert wurden.

Als aussagefähig für die Messung von Lernfortschritten im Lesen wird für das Programm seitens TFA ein Zeitraum von sechs Wochen angegeben. Die Erhebungen für den Prätest wurden daher in der Woche nach Pfingsten 2023 durchgeführt, damit die Posttests eine Woche vor Beginn der Sommerferien in Hessen abgeschlossen werden konnten.

Bei den im Rahmen der Pilotevaluation eingesetzten Testverfahren handelt es sich um wissenschaftlich anerkannte valide und reliable Instrumente, die die geförderten Fähigkeiten unabhängig von den im Tutoring-System selbst implementierten, auf die Bewältigung der konkreten Aufgaben orientierten Einstufungsverfahren messen. Alle mit Elterneinverständnis teilnehmenden Schüler:innen sowohl der Experimental- als auch der Kontrollgruppe wurden mit folgenden Instrumenten getestet:

- ELFE II (Lenhard, Lenhard & Schneider 2022) ist ein diagnostisches Verfahren für Kinder der 1. bis 7. Klasse, mit dessen Hilfe eine umfassende Bewertung des Leseverständnisses auf Wort-, Satz- und Textniveau ermöglicht wird. Die Durchführung erfolgte in Gruppen mit einem Paper-Pencil-Verfahren, die Durchführungszeit betrug insgesamt 20 bis 30 Minuten (davon 13 Minuten reine Bearbeitungszeit).

- Das Salzburger Lesescreening SLS 2-9 (Wimmer & Mayringer 2014) ist ein Screening-Verfahren, mit dem die basalen Lesefertigkeiten von Schülerinnen und Schülern der 2. bis 9. Klasse, ermittelt durch die Lesegeschwindigkeit, erfasst wird. Für die Durchführung als Gruppentest (Paper-Pencil) wurden 15 Minuten inklusive Instruktion und Austeilen der Testhefte veranschlagt, davon beträgt die reine Bearbeitungszeit 3 Minuten.

Für die Durchführung der Tests wurden vorab sechs Testleiter:innen durch das Team der Universität Hamburg geschult. Die Schulung umfasste eine Einführung in das Projekt, die Vorstellung und Einübung der testgemäßen Durchführungsmodalitäten sowie eine Vorstellung der Datenschutzmaßnahmen.

Zur Prüfung der Vergleichbarkeit der Ergebnisse und für die Gewinnung von Variablen für eine statistische Kontrolle von unerwünschten Störgrößen wurde ein Elternfragebogen entwickelt. Darin wird erfragt:

- zuerst in der Familie erworbene Sprache(n) des Kindes: Deutsch – Deutsch und andere Sprache – andere Sprache(n) als Deutsch,
- zu Hause überwiegend gesprochene Sprache: Deutsch – Deutsch und andere Sprache - andere Sprache(n) als Deutsch,
- sozio-ökonomischer Status: Berufe der Eltern, Schulbildung der Eltern,
- Einstellungen zum Lesen,
- kulturelles Kapital (Anzahl der Bücher im Haushalt),
- Migrationshintergrund von Eltern und Kind (geboren in Deutschland).

Die Durchführung der Datenerhebung erfolgte in Räumlichkeiten, die InSL e.V. für die Förderung zur Verfügung stehen. Die Daten wurden an das Evaluationsteam in pseudonymisierter Form übergeben, um zu gewährleisten, dass die Auswertungen in Hamburg ohne Kenntnis der Person durchgeführt wurden. Dazu wurden von InSL e.V. Listen mit den teilnehmenden Schüler:innen geführt, in denen ihre Namen einer für das gesamte Verfahren gültigen Identifikationsnummer zugeordnet wurden. Diese Listen verblieben bis zum Abschluss der Datenanalysen beim Förderverein und wurden anschließend vernichtet. Testhefte und Fragebögen enthielten nur die Identifikationsnummern, keine Namen, Adressen oder andere Informationen, die zur Identifizierung führen könnten. Testergebnisse wurden individuell in pseudonymisierter Form zurückgemeldet, Fragebogenantworten wurden nicht oder nur in tabellarischer aggregierter Form ohne Pseudonyme berichtet.

Zusammengefasst wurde die Wirksamkeitsprüfung wie folgt realisiert:

- Konzeptgemäß wurde ein Prä-Post-Design (Testung vor Beginn und nach Abschluss des Tutorings) mit einer Experimental- und einer Kontrollgruppe durchgeführt.

- Eine angestrebte zufallsgesteuerte Zuweisung konnte seitens des Förderträgers nicht erfolgen.
- Mit Elterneinverständnis nahmen am Prätest 67 und am Posttest 54 geförderte Kinder aus den Klassenstufen 2 und 3 an den Tests teil.
- Die Experimentalgruppe erhielt eine Förderung mit dem digitalen TFA-Programm.
- Die Kontrollgruppe wurde nach konventioneller Methode des InSL e.V. gefördert.
- Zwischen Prä- und Posttest lag eine Förderphase von sechs Wochen.
- Die verwendeten anerkannten Testverfahren ELFE II (Leseverständnis auf Wort-, Satz- und Textebene) und Salzburger Lesescreening SLS 2-9 (basale Lesefertigkeit) messen die Lesefähigkeiten unabhängig von den im Tutoring-System implementierten Einstufungsverfahren.
- Zur Prüfung der Vergleichbarkeit der Ergebnisse und für die Gewinnung von Variablen für eine statistische Kontrolle von unerwünschten Störgrößen wurde ein Elternfragebogen entwickelt und eingesetzt.
- An der schriftlichen Elternbefragung nahmen Eltern von 32 getesteten Kindern teil.

4. Analysestrategie und deskriptive Befunde

Die Auswertungen der erhobenen Daten folgen dem Prinzip eines „fairen Vergleichs“, d.h. sie berücksichtigen potenziell ungleiche Voraussetzungen der teilnehmenden Kinder, die ihre Testleistungen und deren Fortschritte beeinflussen können, indem diese Voraussetzungen statistisch kontrolliert werden und somit keinen Einfluss auf die Resultate nach Testgruppen haben. Systematische Unterschiede zwischen den Teilnehmenden der zu vergleichenden Testgruppen waren aufgrund der kleinen Fallzahlen und der nicht zufälligen Zuweisung zur Experimental- und Kontrollgruppe zu erwarten. Das Untersuchungsdesign hatte aus diesem Grund die Erhebung von Kontrollvariablen im Erhebungsbogen und im Elternfragebogen vorgesehen.

Für die Analysen der Lesefortschritte wurden nur die Daten der 54 Kinder herangezogen, für die Ergebnisse sowohl des Prätests als auch des Posttests vorliegen.

Da nur 32 Elternfragebögen vorlagen, wurden die Antworten der 22 fehlenden Fälle mithilfe von multiplen Imputationen geschätzt, um eine höhere Teststärke zu erzielen. Dafür wurden mittels SPSS mit den im Modell verwendeten Variablen 20 unterschiedliche Datensätze mit Schätzungen für die fehlenden Werte simuliert (Urban et al. 2016). Die für unsere Analysen verwendeten Regressionsmodelle wurden mit allen 20 Datensätzen berechnet, ihre Ergebnisse werden in SPSS nach Rubins Regeln zusammengefasst (Little & Rubin 1987). Die hier berichteten Resultate unserer Analysen beruhen auf diesen gepoolten Schätzergebnissen der multiplen Imputationen.

Im Folgenden werden zunächst die berücksichtigten Testwerte und deren durchschnittliche Verteilung in den Erhebungen vorgestellt. Hierbei handelt es sich um die tatsächlich erzielten Testergebnisse, sie berücksichtigen nicht die unterschiedliche Zusammensetzung der Testgruppen. Für den fairen Vergleich (siehe Kapitel 5) werden multiple Regressionsmodelle berechnet, die das Ergebnis des jeweiligen Posttests in Abhängigkeit zu der Testgruppe, den Resultaten des Prätests und weiteren Kontrollvariablen setzen. Somit wird eine Einschätzung des (Netto-)Fortschritts der Experimental- gegenüber der Kontrollgruppe ermöglicht, da Einflüsse der übrigen in den Modellen berücksichtigten Merkmale und Voraussetzungen der Teilnehmenden statistisch herausgerechnet sind.

4.1. Getestete Lesefähigkeiten

Als Bemessungsgrundlage für die Lesefähigkeiten im Prätest und im Posttest dienen die an den testspezifischen Referenzwerten orientierten Lesequotienten (SLS 2-9) bzw. T-Werte (ELFE II). Der Lesequotient des SLS 2-9 berücksichtigt Klassenstufe und Geschlecht der Kinder sowie die jeweils eingesetzte Testform. Die T-Werte der ELFE II-Subtests berücksichtigen die Klassenstufe und den Schulmonat des laufenden Schuljahrs, in dem der Test stattfand.

Die Testinstrumente sind so konstruiert, dass die Ergebnisse in den Referenzstichproben die folgenden Mittelwerte und Standardabweichungen, d.h. die durchschnittlichen Abweichungen der Werte vom jeweiligen Mittelwert aufweisen:

1. Basale Lesefertigkeit (Salzburger Lese-Screening SLS 2-9): Lesequotient (Mittelwert[MW]=100, Standardabweichung[SD]=15)
2. Wortverständnis (ELFE II, Subtest 1): T-Wert (MW =50, SD =10)
3. Satzverständnis (ELFE II, Subtest 2): T-Wert (MW=50, SD =10)
4. Textverständnis (ELFE II, Subtest 3): T-Wert (MW =50, SD =10)

Zur Beurteilung der Testleistungen sind hier die Zahlenwerte von Lesequotient SLS 2-9 und T-Werten ELFE II für Wortverständnis, Satzverständnis und Textverständnis nach Leistungsgruppen ihrer verbalen Umschreibung gegenübergestellt.

Lesequotient für Basale Lesefertigkeit (Salzburger Lese-Screening SLS 2-9)

≥ 130	sehr gut
120-129	gut
110-119	überdurchschnittlich
90-109	durchschnittlich
80-89	unterdurchschnittlich
70-79	schwach
≤ 69	sehr schwach

(nach Wimmer & Mayringer 2014, 21)

T-Werte für Wort-, Satz- und Textverständnis (ELFE II)

> 70	stark überdurchschnittlich
60-70	überdurchschnittlich
50-60	oberer Normalbereich
40-50	unterer Normalbereich
30-40	unterdurchschnittlich
≤ 30	stark unterdurchschnittlich

(nach Lenhard, Lenhard & Schneider 2022, 73)

4.2. Resultate der Tests, Zusammensetzung der Testgruppen und Folgen für die Regressionsmodelle

Die folgende Tabelle 1 gibt die arithmetischen Mittel der Testergebnisse für Experimental- und Kontrollgruppe wieder. Nicht erwartet wurde, dass sich, verglichen mit den Referenzstichproben, die Mittelwerte bereits im Prätest überwiegend im Normalbereich befinden. Zwar zeigen die Standardabweichungen, dass es teilweise

deutliche Abweichungen von diesen Mittelwerten gibt, dennoch stellt sich die Frage nach den Kriterien der Schulen für ihre Zuweisungspraxis zu Förderangeboten.

Tabelle 1

Arithmetische Mittel und Standardabweichungen der Testergebnisse

Testgruppe	SLS Basale		ELFE II		ELFE II		ELFE II	
	Lesefertigkeit		Wortverständnis		Satzverständnis		Textverständnis	
	-Lesequotienten-		-T-Werte-		-T-Werte-		-T-Werte-	
	Prätest	Posttest	Prätest	Posttest	Prätest	Posttest	Prätest	Posttest
Lesen mit dem Turbo-Team (n=33)	97,6 (15,5)	101,0 (14,4)	54,7 (8,1)	57,8 (9,5)	51,0 (6,9)	54,7 (7,8)	48,9 (8,0)	52,6 (8,4)
Kontrollgruppe (konventionelle Förderung) (n=21)	85,2 (21,5)	87,7 (25,1)	52,0 (11,9)	53,9 (11,0)	45,7 (11,4)	46,1 (10,9)	43,7 (11,8)	43,3 (10,4)

Weiterhin zeigt sich, dass die Kontrollgruppe bereits im Prätest durchweg niedrigere Mittelwerte und höhere Standardabweichungen erzielt hat als die Gruppe, die das Tutoring *Lesen mit dem Turbo-Team* erhalten hat. Die Kontrollgruppe weist somit einen höheren Anteil von Kindern mit schwächeren Leseleistungen auf als die Experimentalgruppe. Diese Unterschiede unterstützen noch einmal die Analysestrategie, die Ergebnisse der Prätests als kontrollierende, unabhängige Variable in den Regressionsmodellen zu verwenden, da so die Effekte des Tutorings unabhängig von den Leistungen des Prätests geschätzt werden können. Zusätzlich ist durch Einbeziehung entsprechender Interaktionseffekte zu überprüfen, ob das bestehende unterschiedliche Leistungsniveau Auswirkungen auf die Effekte des Tutorings hat, ob also gegebenenfalls leistungsstärkere oder leistungsschwächere Kinder in unterschiedlichem Ausmaß davon profitieren.

Die Testgruppen unterscheiden sich in ihrer Zusammensetzung auch nach Geschlecht und Klassenstufe sowie nach der zuerst in der Familie erlernten Sprache (Erstsprache)/vgl. Tabelle 2). So waren in der Tutoring-Gruppe Mädchen, Drittklässler und Kinder mit einer anderen Erstsprache als Deutsch überrepräsentiert, weshalb diese Variablen als Kontrollvariablen berücksichtigt werden müssen. Weniger deutlich unterscheiden sich die Gruppen hinsichtlich des sozio-ökonomischen Status (HISEI¹) der

¹ Ein häufig verwendetes Maß für den sozio-ökonomischen Status ist der International Socio-Economic Index of Occupational Status (ISEI), der den Status auf Basis der ausgeübten bzw. zuletzt ausgeübten Berufe zuweist (Ganzeboom & Treiman 2003). Unsere Klassifizierung erfolgte anhand der Berufsangaben von Vater und Mutter in den Elternfragebögen mittels des ISEI-08 (Ganzeboom 2010). Als Wert für den Status der Familie wurde der jeweils status-höhere Beruf von Vater oder Mutter angenommen (HISEI).

Familien. Da diese Variable aber auch mit sprachlichen Leistungen korreliert, soll sie als strukturelles Merkmal ebenfalls kontrolliert werden.²

Tabelle 2

Kontrollvariablen nach Testgruppe

		Experimental- gruppe <i>Lesen mit dem Turbo-Team</i>	Kontroll- gruppe <i>konventionelle Förderung</i>
Geschlecht (n)	weiblich	23	8
	männlich	10	13
Klassenstufe (n)	2	13	17
	3	20	4
Erstsprache* (n)	Deutsch	5	5
	Deutsch und and. Sprache	5	6
	andere Sprache	23	10
Sozio.-ökonom. Status* (Mittelwert/Standardfehler)	HISEI	45,4/10,1	50,2/19,6

* multiple Imputation fehlender Werte

² Der Elternfragebogen enthält deutlich mehr Variablen. Aufgrund der kleinen Fallzahl kann nur eine begrenzte Anzahl von unabhängigen Variablen in das Modell aufgenommen werden. Wir haben hier die aus Verteilungsdifferenzen notwendigen und gleichzeitig theoretisch sinnvollen Variablen berücksichtigt, die darüber hinaus in Bezug auf die Erklärung der abhängigen Variablen wenig gemeinsame Varianz miteinander teilen und somit Probleme der Multikollinearität vermeiden.

5. Die Resultate im fairen Vergleich

Für eine faire Beurteilung der Testergebnisse wurden die Resultate der Posttests auf Basis der Prätestergebnisse, der Teilnahme am TFA-Tutoring unter Hinzufügen der genannten Kontrollvariablen mit multiplen Regressionsmodellen geschätzt. Für jeden Test bzw. Subtest wurde ein eigenes Regressionsmodell mit denselben unabhängigen Variablen gerechnet.

Im Vorfeld waren mögliche Einflüsse weiterer unabhängiger Variablen geprüft worden, die aber zu keinen nennenswerten Unterschieden oder zusätzlichen Erkenntnissen beitrugen und verworfen wurden, um die Modelle bei kleiner Fallzahl sparsam zu halten.

Die erwähnte Prüfung der Frage, inwieweit Effekte der Förderung abhängig vom zuvor bestehenden Leistungsniveau der Kinder waren, wurde ebenfalls im Vorfeld vorgenommen. Dazu wurden in die finalen Regressionsmodelle Interaktionseffekte von TFA-Förderung und Prätestergebnis einbezogen. Dabei zeigte sich, dass in den Analysen für alle vier abhängigen Variablen keine signifikanten Interaktionseffekte auftraten. Der p -Wert für die Signifikanz betrug bei der Basalen Lesefähigkeit (SLS) $p=0,615$, beim Wortverständnis $p=0,144$, beim Satzverständnis $p=0,942$ und beim Textverständnis $p=0,497$. Da die Haupteffekte bei Berücksichtigung von Interaktionseffekten schwer zu interpretieren (Mayerl & Urban 2019) und letztere für unsere Analysen nicht bedeutsam sind, wurden die Interaktionen in der Folge nicht mehr einbezogen.

Die multiple Regression (Fahrmeir et al. 2009) ist ein multivariates statistisches Verfahren, das die Effekte der unabhängigen Variablen oder Prädiktoren auf die abhängige Variable simultan schätzt. Dabei werden die Beiträge, die jede einzelne unabhängige Variable als Erklärung zu der Varianz der abhängigen Variable leistet, „bereinigt“ um die Anteile der Varianz, die auf Korrelationen der unabhängigen Variablen untereinander zurückgehen. Man kann also von Netto-Beiträgen jeder im Modell berücksichtigten unabhängigen Variable sprechen. In unseren Modellen sollten die Testergebnisse der Posttests in Basaler Lesefähigkeit und im Wort-, Satz- und Textverständnis erklärt werden durch die Ergebnisse der Prätests, der besonderen Förderung des TFA-Tutorings sowie durch das Geschlecht der Kinder, ihre Klassenstufe, ihre zuerst in der Familie erworbene(n) Sprache(n) und den sozio-ökonomischen Status ihrer Familie.

Dabei interessiert uns *de facto* für die Evaluation der Wirksamkeit ausschließlich der eigenständige, um die Einflüsse der anderen unabhängigen Variablen bereinigte Beitrag, den das Tutoring für den Lernfortschritt³ gegenüber der konventionellen Förderung leistet. Beide Förderungen sind in einer einzigen Variablen repräsentiert, in

³ Dadurch, dass in allen (Kovarianz-)Modellen die Posttestresultate in Abhängigkeit von den jeweiligen Prätest-Ergebnissen berechnet werden, repräsentieren die verbleibenden zu erklärenden Varianzen der abhängigen Variablen die Veränderungen, die zwischen beiden Messzeitpunkten stattgefunden haben (Möbus & Nagl 1983), hier also Fort- oder Rückschritte in den Testergebnissen.

der die Beteiligung am TFA-Tutoring den Wert 1 und die Zugehörigkeit zur Kontrollgruppe den Wert 0 erhält. Der Netto-Effekt des Tutorings zeigt sich dann in der Höhe des nicht standardisierten Regressionskoeffizienten. Dieser gibt an, um wie viele Testpunkte das Posttestergebnis gegenüber dem der Kontrollgruppe durchschnittlich höher ausfällt, wenn ein sechswöchiges TFA-Tutoring stattgefunden hat – unabhängig von Prätestergebnis, Geschlecht, Klassenstufe, Erstsprache und sozio-ökonomischem Status der Familie. Um einen hinreichenden Hinweis auf die Wirksamkeit des Tutorings geben zu können, sollte dieser Effekt gegen den Zufall abgesichert sein, wobei wir konventionell eine Irrtumswahrscheinlichkeit von höchstens 5 Prozent voraussetzen.

5.1. Ergebnisse

Die Ergebnisse der multiplen Regressionsanalysen sind in Tabelle 3 zusammengestellt. Aufgrund der unterschiedlichen Skalierung ist der Wert des Lesequotienten des Salzburger Lese-Screenings (M 100, SD 15) nicht unmittelbar mit den T-Werten der ELFE II-Subtests (MW 50, SD 10) vergleichbar.

Gegenüber der Kontrollgruppe, die eine konventionelle Leseförderung erhalten hat, haben die Kinder mit einem TFA-Tutoring nach sechs Wochen durchschnittlich einen um 2,3 Punkte höheren Lesequotienten erreicht, wenn zusätzlich Geschlecht, Klassenstufe, Erstsprache und HISEI kontrolliert werden. Der Unterschied zwischen den Gruppen ist allerdings klein und mit $p=0,456$ statistisch nicht signifikant, sodass er nicht gegen den Zufall abgesichert werden kann.

In Bezug auf das Wortverständnis ergibt sich unter Konstanthaltung der übrigen Prädiktoren für die Tutoring-Gruppe nach sechs Wochen Förderung im Mittel ein um 1,72 Punkte höherer T-Wert als bei der Kontrollgruppe. Auch dieser höhere Zuwachs ist statistisch nicht signifikant.

Deutlich größere und signifikante mittlere Unterschiede zwischen den beiden Fördergruppen ergeben sich hingegen bei den Zuwächsen im Satzverständnis (4,53 Punkte, $p=0,016$) und im Textverständnis (5,17 Punkte, $p=0,015$). Die T-Werte sind damit bei den von uns getesteten Kindern mit sechswöchigem Tutoring durchschnittlich um etwa eine halbe Standardabweichung höher als bei den konventionell geförderten Kindern.

Tabelle 3
Ergebnisse der multiplen Regressionsmodelle

abhängige Variablen	SLS Posttest Lesequotient			Wortverständnis Posttest T-Wert			Satzverständnis Posttest T-Wert			Textverständnis Posttest T-Wert		
	nicht standard. Regressionskoeffizient <i>b</i>	Standardfehler	Signifikanz <i>p</i>	nicht standard. Regressionskoeffizient <i>b</i>	Standardfehler	Signifikanz <i>p</i>	nicht standard. Regressionskoeffizient <i>b</i>	Standardfehler	Signifikanz <i>p</i>	nicht standard. Regressionskoeffizient <i>b</i>	Standardfehler	Signifikanz <i>p</i>
Konstante	7,45	12,21	0,543	0,98	5,56	0,861	6,91	6,57	0,293	12,24	6,51	0,060
TFA (vs. Kontrollgruppe)	2,30	3,09	0,456	1,72	1,49	0,249	4,53	1,87	0,016	5,17	2,13	0,015
Prätest	0,86	0,09	0,000	0,95	0,07	0,000	0,85	0,10	< 0,001	0,72	0,10	< 0,001
weiblich (vs. männlich)	0,18	3,32	0,957	1,13	1,48	0,445	0,08	1,76	0,965	-1,44	1,98	0,467
Klasse 3 (vs. Klasse 2)	1,31	2,86	0,647	1,30	1,38	0,345	-0,78	1,62	0,630	-1,71	1,89	0,368
Erstsprache (Dt. / Dt.+and./ And. Spr.)	0,74	2,61	0,777	-0,77	1,15	0,506	0,39	1,42	0,783	2,19	1,57	0,166
HISEI	0,06	0,05	0,267	0,03	,02	0,176	0,03	0,03	0,368	-0,01	0,03	0,794

5.2. Diskussion

Ein Vergleich der TFA-Gruppe mit der Kontrollgruppe hatte ergeben, dass sie sich in ihrer Zusammensetzung unterscheiden und vor Beginn der Förderung Leistungsunterschiede aufweisen. Dadurch wurde es notwendig, die Unterschiede statistisch auszugleichen.

Im fairen Vergleich, d.h. unter statistischer Kontrolle vom Prätestergebnis, von Geschlecht, Klassenstufe, der zuerst in der Familie erworbenen Sprache und dem sozioökonomischen Status der Familie, weisen die Kinder, die eine sechswöchige Förderung mit *Lesen mit dem Turbo-Team* erhalten haben, in allen getesteten Lesefähigkeiten durchschnittlich bessere Resultate auf als die Kinder, die im gleichen Zeitraum eine konventionelle Förderung des InSL e.V. erhalten haben.

Die geringen und nicht signifikanten Unterschiede in den Basalen Lesefertigkeiten und im Wortverständnis sind allerdings vernachlässigbar. Hier ist von vergleichbaren Fortschritten mit dem Tutoring wie auch mit der Förderung in der Kontrollgruppe auszugehen.

Signifikant und von der Effektstärke her bemerkenswert sind die Unterschiede zwischen beiden Gruppen in der mittleren Zunahme des Wort- und des Textverständnisses, also bei anspruchsvolleren Lesekompetenzen, die das Erfassen von inhaltlichen Zusammenhängen betreffen. Dieses Resultat deckt sich in der Tendenz mit den Ergebnissen der Evaluationsstudie von Ross et al., die für das amerikanische Tutoring-Programm ebenfalls einen substantiellen Einfluss bei komplexeren Fähigkeiten (word synthesis und text comprehension) nachweisen konnte (Ross et al. 2017, 10). Die Netto-Unterschiede zwischen beiden Gruppen liegen jeweils im Bereich einer halben Standardabweichung der T-Wert-Skalen für diese Lesekompetenzen – und damit in der Größenordnung des mittleren Lernzuwachses im Leseverstehen innerhalb eines Schuljahres¹, wenn keine weitere Förderung erfolgt (vgl. Hornberg et al. 2007). Dieser Unterschied nach einer nur sechswöchigen Förderung ist beachtlich, allerdings wurde nur ein unmittelbarer, kurzzeitiger Effekt des Trainings erfasst, keine längerfristige Wirksamkeit des Tutorings. Außerdem ist bei den kleinen Fallzahlen eine Generalisierbarkeit der Effektstärke über die getestete Stichprobe hinaus problematisch. Dennoch ist das Ergebnis auch wegen der übereinstimmenden Tendenzen mit der Evaluation für die amerikanische Originalversion ein deutlicher Hinweis auf das Potenzial des TFA-Programms. Dabei ist noch einmal hervorzuheben, dass das Tutoring nicht gegen eine Non-Intervention, sondern gegen eine konventionelle Leseförderung getestet wurde.

Geprüft wurde auch, ob Interaktionseffekte, die eine Abhängigkeit der Wirkung des Tutoring vom vorher bestehenden Kompetenzniveau beschreiben, die Ergebnisse beeinflussen. Ein Interaktionseffekt könnte z.B. darin bestehen, dass das Tutoring erst

¹ Diese Angabe bezieht sich auf Schüler:innen am Ende der Grundschule (Klasse 4).

wirkt, wenn ein:e Schüler:in bereits ein bestimmtes Niveau von Lesefähigkeit erreicht hat, oder auch darin, dass es bei Leistungsschwächeren besser als bei Fortgeschrittenen wirkt. Für die von uns erhobenen Daten konnte kein signifikanter Interaktionseffekt nachgewiesen werden, sodass wir für das Leistungsspektrum unserer Stichprobe von einem vom Ausgangsniveau unabhängigen Effekt ausgehen konnten.

Die Untersuchung ist als Teil einer Pilot-Evaluation konzipiert und durchgeführt worden und entspricht demgemäß nicht den Anforderungen einer regulären Wirkungsevaluation der Maßnahme. Die Ergebnisse stehen insgesamt unter dem Vorbehalt der kleinen Fallzahl und der Tatsache, dass keine randomisierte Zuweisung der Schüler:innen zur Maßnahme- und zur Kontrollgruppe stattfinden konnte. Letzteres schlägt sich unter anderem in den genannten unterschiedlichen mittleren Leistungsniveaus der beiden Gruppen nieder, da die Kontrollgruppe in allen Tests in Erst- und Zweitmessung durchschnittlich schlechter abschnitt als die Turbo-Team-Gruppe. Durch die Berücksichtigung von Kontrollvariablen wurde den Gruppenunterschieden (teilweise) Rechnung getragen, diese *ex-post*-Korrektur kann aber ein randomisiertes Verfahren nicht ersetzen. Eine entsprechend große Stichprobe böte darüber hinaus zusätzliche Möglichkeiten, die Wirksamkeit des Tutorings nach unterschiedlichen Subgruppen zu untersuchen.

6. Durchführung und Resultate der Tutor:innen-Befragung

Die Wirksamkeit einer Maßnahme ist entscheidende Grundvoraussetzung für die Empfehlung ihrer Anwendung. Die Pilot-Evaluation hat für das deutsche Tutoring-Programm *Lesen mit dem Turbo-Team* ermutigende Hinweise auf wirksame Fördereffekte ergeben, wie sie bereits für das amerikanische Vorbild nachgewiesen wurden. Die sich in Testergebnissen niederschlagenden Effekte sind dennoch nur die eine Seite der Medaille. Auf der anderen Seite stehen die Erfahrungen und Eindrücke derjenigen, die das Tutoring anwenden, und derjenigen, die damit gefördert werden. Hoher Aufwand, mangelnde Akzeptanz oder Motivation, monetäre und Zeit-Kosten, Nichteignung für bestimmte Zielgruppen und mehr können auch bei nachgewiesenen Erfolgen einer Maßnahme zu einer negativen Einschätzung durch die Anwender:innen führen, die den eigenen Aufwand und die Kosten in Relation zum Nutzen setzen. Das Evaluationsdesign hatte aus diesem Grund eine Befragung der mit dem TFA-Programm praktizierenden Tutor:innen vorgesehen. Eine Befragung der teilnehmenden Kinder war nicht geplant, jedoch sollten im Fragebogen für Tutor:innen deren Wahrnehmungen zur Akzeptanz, Schwierigkeit oder Teilnahmemotivation der Kinder berücksichtigt werden.

Die Befragung sollte online zum Jahreswechsel 2023/2024 erfolgen und alle bis dahin geschulten und praktizierenden Tutor:innen erfassen. Damit ging der anvisierte Teilnehmendenkreis von der Anzahl und der individuellen Dauer der Praxiserfahrung her sowie regional über die Personengruppe hinaus, die das Tutoring mit den an der tentativen Wirksamkeitsprüfung beteiligten Kindern durchgeführt haben.

Die E-Mail-Adressen der in Frage kommenden Personen wurden von TFA zur Verfügung gestellt. Im Vorfeld war die Genehmigung zur Bereitstellung der Adressen von den beteiligten Institutionen (Fördereinrichtungen, Schulen) eingeholt worden. Es standen 54 Adressen zur Verfügung. Nicht berücksichtigt wurden dabei die Adressen von Leitungspersonen und geschulten Tutor:innen, von denen an den beteiligten Institutionen bekannt war, dass sie (noch) nicht mit dem Programm *Lesen mit dem Turbo-Team* gearbeitet hatten.

Die Einladung zur Befragung an alle 54 Adressen wurde am 19. 12. 2023 mit einem persönlichen Link zum Online-Fragebogen verschickt. Die Erhebungsphase endete am 9. 1. 2024. Die Beantwortung der Fragen erfolgte anonym, eine individuelle ID wurde vom Befragungssystem vergeben. Zwei Erinnerungs-E-Mails an die Personen, die zu diesen Zeitpunkten noch nicht teilgenommen oder ihre Teilnahme nicht abgelehnt hatten, wurden vom System automatisch verschickt.

Der Rücklauf der Befragung ist in Tabelle 4 zusammengestellt. Aus Benachrichtigungen von vier Angeschriebenen ergab sich, dass sie bisher keine Praxiserfahrungen mit dem TFA-Programm machen konnten. Diese Personen wurden aus der Befragtenliste entfernt, sodass sich eine Bruttostichprobe von 50 Personen ergibt. Von diesen lehnten vier eine Teilnahme explizit ab, weitere vier reagierten nicht auf die Einladung. Vier

Personen starteten die Umfrage, brachen sie aber nach der ersten Frage ab. Gültige Teilnahmen lagen schließlich von 38 Personen vor. Das entspricht einer Rücklaufquote von 76 Prozent der bereinigten Bruttostichprobe. Zwar gibt es keine gültigen Zielwerte für Online-Befragungen mit einer geschlossenen Befragtengruppe, jedoch werden z.B. für Online-Patientenbefragungen Werte von 60 Prozent (IQTIG 2021) oder für Mitarbeiterbefragungen „typische Rücklaufquoten“ von 50 bis 80 Prozent (Theobald 2017) genannt. Damit kann in unserem Fall von einem guten bis sehr guten Rücklauf ausgegangen werden.

Tabelle 4

Brutto- und Nettostichprobe der Tutor:innen-Befragung

	<i>n</i>	<i>%</i>
angeschriebene Personen	54	
Fehladressen (keine Praxis mit TFA)	4	
bereinigte Bruttostichprobe	50	100
keine Reaktion	4	
Ablehnungen	4	
	42	84,0
nach Frage 1 abgebrochen	4	
gültige Teilnehmende	38	76,0

6.1. Teilnehmende an der Befragung

Die Befragtengruppe setzt sich wie folgt zusammen: ein Drittel sind Tutor:innen, die zum Befragungszeitraum bereits seit mehr als drei Monaten mit dem Programm arbeiteten; knapp die Hälfte konnte auf Tutoring-Erfahrungen zwischen einem und drei Monaten zurückblicken und bei etwa einem Fünftel handelt es sich um Einsteiger:innen mit höchstens einem Monat Praxis in der TFA-Förderung (Abb. 1).

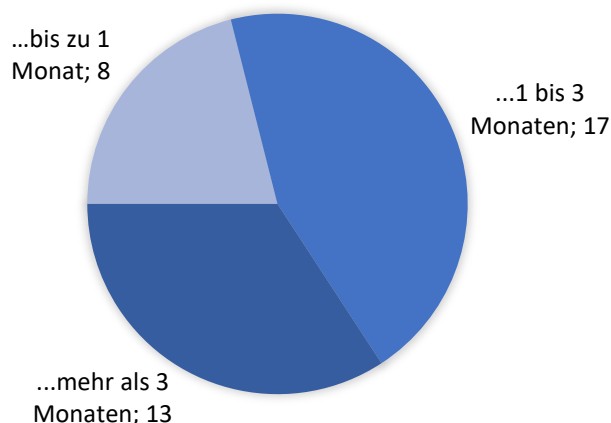


Abbildung 1: Seit wann arbeiten Sie als Tutor(in) mit „Lesen mit dem Turbo-Team“?

Generelle Erfahrungen mit der Förderung von Lesefähigkeiten (Abb. 2) seit mehr als einem Jahr hatte ein Drittel der Befragten, ein Fünftel war zwischen drei Monaten und einem Jahr mit der Leseförderung befasst und knapp die Hälfte war mit bis zu drei Monaten erst vor Kurzem in dieses Praxisfeld eingestiegen. Letztere sind

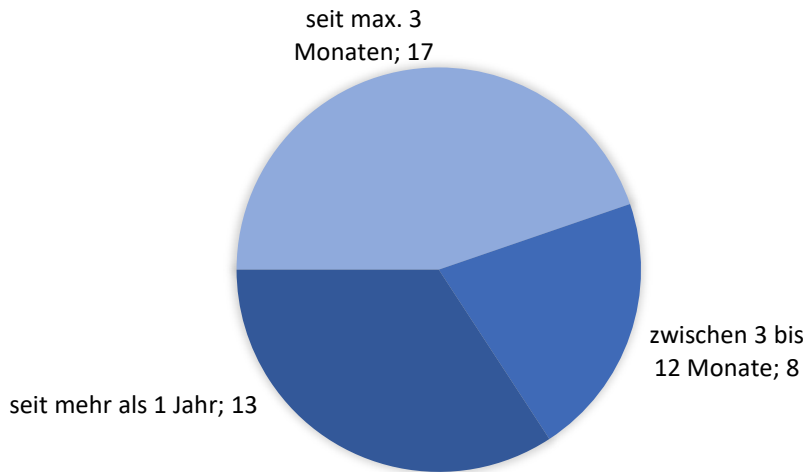


Abbildung 2: Seit wann sind Sie schon insgesamt im Rahmen der Leseförderung tätig?

bis auf eine Ausnahme erst mit der TFA-Förderung im Bereich der Leseförderung tätig geworden (Tabelle 5). Für knapp die Hälfte der Befragten waren Vergleichsmöglichkeiten mit der Praxis anderer Ansätze der Leseförderung somit nur begrenzt vorhanden. Für mehr als die Hälfte lagen aber mehr oder weniger Erfahrungen damit vor.

Tabelle 5

Erfahrung mit Leseförderung generell und mit „Lesen mit dem Turbo-Team“

In der Leseförderung	Förderung mit <i>Lesen mit dem Turbo-Team</i>			Gesamt
	bis zu 1 Monat	1 bis 3 Monate	mehr als 3 Monate	
innerhalb der letzten 3 Monate	4	12	1	17
3 bis 12 Monate	1	1	6	8
mehr als 1 Jahr	3	4	6	13
Gesamt	8	17	13	38

Die Frage nach den Klassenstufen, denen die von den Tutor:innen mit TFA geförderten Kinder angehören¹, zeigte, dass das Programm zwar auf den Lerninhalten der Klassenstufen 1 bis 4 beruht, aber bis zur 8. Klassenstufe eingesetzt wurde (Abbildung 3). Zwar kam das Tutoring überwiegend in den Klassenstufen 2 bis 4 und unmittelbar nach dem Übergang in die Sekundarstufe zur Anwendung, doch zeigt sich, dass 10 der

¹ Die hier beurteilten Kinder sind nicht identisch mit der Stichprobe derjenigen, die in die oben vorgestellte Evaluation einbezogen wurden.

befragten Tutor:innen das Programm mit Schüler:innen der Klassen 6 bis 8 einsetzen. Dieser Umstand ist für die Bewertung der nachfolgenden Analysen zu beachten.

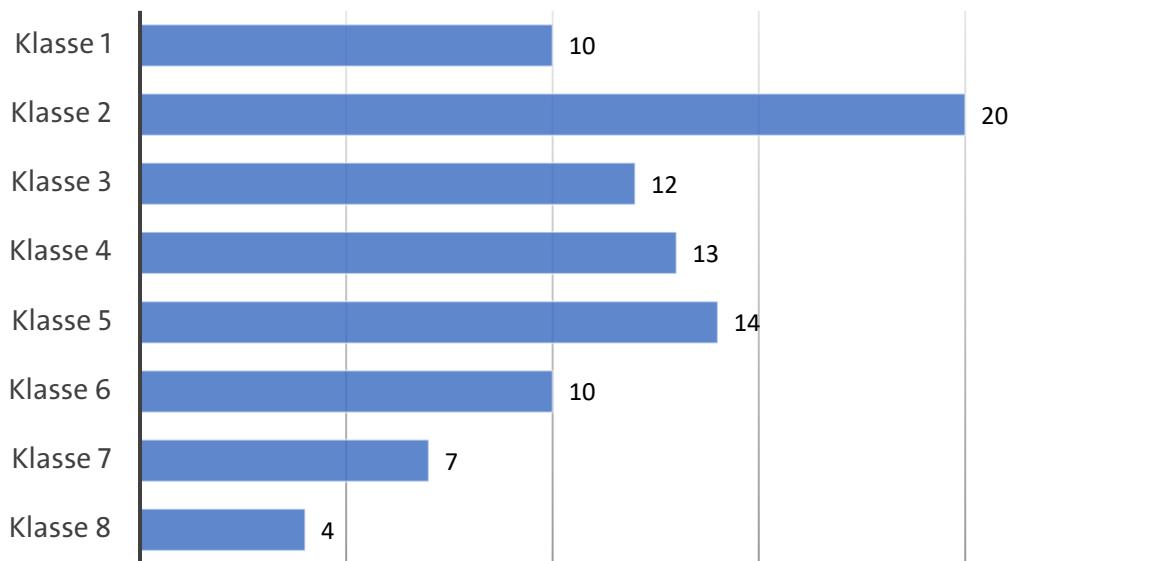


Abbildung 3: Anzahl der Tutor:innen mit Schüler:innen der Klassenstufen (mit Mehrfachnennungen)

6.2. Einschätzungen und Erfahrungen der Tutor:innen

Bei der Untersuchung von Akzeptanz und Wirksamkeitsvermutung haben wir uns an Studien zu entsprechenden Fragestellungen angelehnt, die mit Lehrkräften durchgeführt wurden. Für diesen Zusammenhang wird davon ausgegangen, dass evidenzbasierte Leseförderkonzepte erst dann ihr volles Wirkungspotential im Unterricht entfalten können, „wenn Verhalten und Einstellungen [der Lehrenden] miteinander im Einklang stehen“ (Meudt et al. 2020, 141). Akzeptanz und Machbarkeit stellen demnach – neben anderen – wichtige Bedingungen für den Transfer in die schulische Praxis dar. Für die Messung entsprechender bedeutender Erfolgskriterien im Transferprozess haben Meudt u.a. fünf Kurzskalen entwickelt, die sich an Lehrkräfte richten, die im Bereich der Leseförderung tätig sind.

Auch im Kontext von tutorieller Förderung sollten die Einstellungen der praktizierenden Personen eine grundlegende Voraussetzung für die erfolgreiche Umsetzung eines Konzepts sein. Dennoch mussten die vorliegenden Skalen adaptiert werden, da die Adressat:innen die Leseförderung zumeist nicht unmittelbar im schulischen Kontext und im regulären Unterricht praktizieren. Daher sind drei der Kurzskalen, die sich auf Machbarkeit im Unterricht, Wiedergabetreue spezifischer Tools im Unterricht sowie Kooperation im Kollegium beziehen, nicht auf den Kontext TFA übertragbar. Eine Kurzskala für „Akzeptanz“, die der Domäne der Überzeugungen und Einstellungen der Tutorin oder des Tutors zuzuordnen ist, sowie eine Skala zu „Lernerfolg“, die auf die Konsequenzen i.S.v. feststellbaren Ergebnissen der Maßnahme

für Schüler:innen gerichtet ist, ließen sich hingegen mit Modifikationen einzelner Items für die Befragung der Tutor:innen verwenden.

6.2.1. Einstellung zu TFA und Einschätzung der Wirksamkeit

Akzeptanz: Die Originalskala (Meudt et al. 2021) besteht aus zehn Items. Die darin formulierten Verweise auf das evaluierte Lese-Tool wurden in unserer Befragung jeweils durch „Lesen mit dem Turbo-Team“ bzw. „diese(r) Förderung“ ersetzt. Zwei Items wurden nicht verwendet, da sie sich als nicht passend auf den Kontext TFA erwiesen.¹ Für unsere Befragung besteht die modifizierte, eindimensionale Skala somit aus acht Items. Sie weist mit einem Cronbachs α von 0,91 eine exzellente Reliabilität auf (Tabelle 6).

Jedes Item wurde von den Befragten auf einer siebenstufigen Likert-Skala eingestuft: 1 = stimme gar nicht zu, 2 = stimme nicht zu, 3 = stimme eher nicht zu, 4 = unentschieden, 5 = stimme eher zu, 6 = stimme zu, 7 = stimme vollkommen zu.

Die Mittelwerte (M) für die *Items* liegen durchweg über 5 (stimme eher zu) bei jeweils einer Standardabweichung (SD) von ca. ± 1 . Das Item „Ich kenne das theoretische Konzept hinter dieser Förderung“ erfährt dabei die niedrigste Zustimmung bei relativ hoher Standardabweichung. Hierin zeigt sich ein Nachholbedarf an entsprechender Information bei einigen der Tutor:innen.

Die *Gesamtskala* weist für die generelle Akzeptanz von TFA seitens der Tutor:innen einen Mittelwert von 5,82 bei einer Standardabweichung von 0,82 auf. Somit trifft das Tutoring durchschnittlich auf eine hohe Akzeptanz zwischen den Beurteilungen „stimme eher zu“ und „stimme zu“.

Die Akzeptanz ist mit einem Mittelwert von 6,18 (SD 0,868) bei erfahrenen Tutor:innen, die bereits seit mehr als einem Jahr in der Leseförderung tätig waren, mit $p=0,034$ signifikant höher als bei den Personen, die erstmalig mit TFA Erfahrungen in der Leseförderung sammeln konnten, d.h. innerhalb der letzten drei Monate vor der Befragung (M 5,43 / SD 0,722).² Dieses Ergebnis spiegelt offenbar eine zurückhaltendere Bewertung durch weniger erfahrene Förderkräfte wider, die noch

¹ Es handelt sich um das Item „Die Lesemappe (bzw. Lesen mit dem Turbo-Team) lässt sich problemlos in den Unterrichtsablauf integrieren“, das inhaltlich nicht mit TFA kompatibel ist, da die Plattform nicht in den Unterrichtsablauf eingebunden ist. Das zweite nicht verwendete Item „Das Arbeiten mit ‚Lesen mit dem Turbo-Team‘ ist im regelmäßigen Gebrauch unkompliziert“ lud in unseren Daten nicht auf derselben Komponente wie die übrigen Items und wurde daher aus der Skala ausgeschlossen.

² Die Mittelwerte für die Akzeptanz-Skala der drei Gruppen (n=38) mit Erfahrung in der Leseförderung (1) „[...] innerhalb der letzten 3 Monate [...] erste Tätigkeit im Rahmen der Leseförderung“, (2) „[...] zwischen drei Monaten und einem Jahr [...]“ sowie (3) „[...] seit mehr als einem Jahr [...]“ wurden mit einer univariaten Varianzanalyse (ANOVA) auf signifikante Unterschiede geprüft. Nach der Levene-Statistik ist von Varianzhomogenität auszugehen ($p=0,417$). Der F -Wert von 4,064 ($df=2$) ist signifikant ($p=0,026$), nach dem Bonferroni Post-Hoc-Test besteht der Unterschied zwischen den Gruppen (1) und (3) ($p=0,034$).

keine Vergleichsmöglichkeiten mit anderen Methoden hatten. Gleichzeitig spricht es für das Tutoring, dass die Akzeptanz bei erfahrenen Kräften besonders hoch ist.

Tabelle 6

Skala Akzeptanz nach Meudt et al. 2020, Hauptkomponentenanalyse und Itemanalyse, finale Lösung mit 8 Items (N=38)

<i>Die folgenden Aussagen beziehen sich auf Ihre Einschätzungen zur Förderung mit „Lesen mit dem Turbo-Team“. Bitte geben Sie bei jeder Aussage an, inwieweit diese Ihrer Meinung nach zutrifft.</i>				
Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		0,880		
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	181,997		
	df	28		
	Signifikanz nach Bartlett	<0,001		
Erklärte Gesamtvarianz		63,0%		
Item*	Komponentenladung	M Mittelwert	SD Standardabw.	r_{it} Trennschärfe
Mir gefällt die Vorgehensweise bei „Lesen mit dem Turbo-Team“.	0,791	5,82	0,955	0,711
Ich kenne das theoretische Konzept hinter dieser Förderung*.	0,618	5,47	1,330	0,532
Viele Schüler(innen) haben Probleme beim Lesen, weshalb der Einsatz dieser Förderung erforderlich ist.	0,717	6,24	0,820	0,628
Insgesamt erachte ich den Einsatz dieser Förderung als lohnenswert.	0,886	6,08	1,100	0,825
Diese Förderung unterstützt die Schüler(innen) beim Erreichen der** vorgesehenen Lernziele.	0,858	5,63	0,970	0,807
Die meisten Schüler(innen) profitieren vom Einsatz dieser Förderung.	0,878	5,87	1,044	0,821
Ich möchte die Qualität meines Förderunterrichts durch „Lesen mit dem Turbo-Team“ steigern.	0,755	5,87	0,906	0,671
Ich bin dafür, „Lesen mit dem Turbo-Team“ regelmäßig im Förderunterricht einzusetzen.	0,810	5,61	1,220	0,739
Gesamtskala Cronbachs α		0,91		

* Die in den Items der Originalskala verwendeten Bezüge auf das Tool „Lesemappe“ wurden in der TFA-Befragung durch „Lesen mit dem Turbo-Team“ bzw. „diese(r) Förderung“ ersetzt.

** Im Originalitem: „im Lehrplan vorgesehenen Lernziele“.

Wirksamkeitseinschätzung: Diese Skala beruht auf der aus fünf Items bestehenden Kurzskala „Lernerfolg der Schüler*innen“ (Meudt et al. 2020, 151). Auch hier mussten an den Items Modifikationen in den Formulierungen vorgenommen werden, da sich die Originalskala auf spezifische Tools („Lesemappe“) und Lernziele („Lesestrategie“) bezieht, die nicht auf TFA zutreffen. Stattdessen beziehen sich die reformulierten Antwortvorgaben wieder auf „Lesen mit dem Turbo-Team“ sowie auf die generellen

Lesefähigkeiten „allgemeine Lesekompetenz“, „Lese Flüssigkeit“ und „Leseverständnis“. Die Antwort erfolgte wie bei der Akzeptanz-Skala auf einer siebenstufigen Likert-Skala von 1 („stimme gar nicht zu“ bis 7 „stimme vollkommen zu“).

Tabelle 7 fasst die für die Skala berechneten Ergebnisse der Hauptkomponenten- und der Itemanalyse zusammen. Alle fünf Items ergeben eine eindimensionale Skala mit einem mit einem guten Cronbachs α -Wert von 0,87.

Tabelle 7

Skala Wirksamkeitseinschätzung nach Meudt et al. 2020, Hauptkomponentenanalyse und Itemanalyse, mit 5 Items (N=37)

Die folgenden Aussagen beziehen sich auf die Wirksamkeit von „Lesen mit dem Turbo-Team“. Bitte geben Sie bei jeder Aussage an, inwieweit diese Ihrer Meinung nach zutrifft.				
Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		0,793		
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	95,231		
	df	10		
	Signifikanz nach Bartlett	<0,001		
Erklärte Gesamtvarianz		65,9%		
Item*	Komponentenladung	M Mittelwert	SD Standardabw.	r _{it} Trennschärfe
„Lesen mit dem Turbo-Team“ wirkt motivierend auf meine Schüler(innen).	0,599	5,46	0,960	0,469
Meine Schüler(innen) lernen durch „Lesen mit dem Turbo-Team“ viel dazu.	0,856	5,38	1,037	0,763
Die allgemeine Lesekompetenz meiner Schüler(innen) verbessert sich durch „Lesen mit dem Turbo-Team“ sehr.	0,850	5,41	0,927	0,727
Die Lese Flüssigkeit meiner Schüler(innen) verbessert sich durch „Lesen mit dem Turbo-Team“ sehr.	0,840	5,49	0,870	0,709
Das Leseverständnis meiner Schüler(innen) verbessert sich durch „Lesen mit dem Turbo-Team“ sehr.	0,880	4,89	1,022	0,788
Gesamtskala Cronbachs α		0,87		

* Die in den Items der Originalskala verwendeten Bezüge auf das Tool „Lesemappe“ wurden in der TFA-Befragung durch „Lesen mit dem Turbo-Team“ ersetzt.

Die Mittelwerte (M) für die Items liegen mit einer Ausnahme auch bei dieser Skala über 5 (stimme eher zu) bei jeweils einer Standardabweichung (SD) von ca. ± 1 . Das Item „Das Leseverständnis meiner Schüler(innen) verbessert sich durch ‚Lesen mit dem Turbo-Team‘ sehr“ erfährt die niedrigste Zustimmung und erreicht nicht ganz den Mittelwert von 5. Motivation und Lernzuwachs bei der allgemeinen Lesekompetenz

sowie der Leseflüssigkeit werden also von den Tutor:innen positiv wahrgenommen. Etwas zurückhaltender sind die Einschätzungen bezüglich einer deutlichen Verbesserung des Leseverständnisses. Hier entspricht der Eindruck der Befragten nicht ganz den oben berichteten tentativen Ergebnissen der Lesetests, bei denen die Wirksamkeitsprüfung auf Satz- und Textebene besonders hohe Zuwächse zeigt. Die Lesetests wurden allerdings nur mit Kindern der Klassenstufen 2 und 3 durchgeführt.

Die Gesamtskala zeigt mit einem Mittelwert von 5,32 und einer Standardabweichung von 0,778 eine positive Einschätzung der Wirksamkeit des Tutorings zwischen „stimme eher zu“ und „stimme zu“.

Wie bei der Akzeptanz, zeigen sich auch hier signifikante Unterschiede zwischen den Einschätzungen der Einsteiger:innen und der in der Leseförderung erfahrenen Kräfte: Bei ersteren liegt der Mittelwert bei 4,94 (SD=0,607), bei letzteren bei 5,62 (SD=0,737).¹ Auch in der Wirksamkeitseinschätzung sehen wir also vorsichtiger (aber nicht negative) Bewertungen der wenig erfahrenen und positivere der erfahreneren Personen. Hervorzuheben ist auch hier die bessere Bewertung des Lernerfolgs durch Tutor:innen, die Vergleiche mit anderen Fördermethoden ziehen können.

6.2.2. Praxiserfahrungen, Durchführbarkeit und wahrgenommene Eignung von TFA

Zum Thema der Machbarkeit wurde in der Befragung nach Vor- und Nachteilen des Tutorings und seiner Rahmenbedingungen gefragt, denen die Tutor:innen in der praktischen Anwendung begegnet sind. Die Fragen richteten sich auf die Themen „Schulung und Begleitung der Tutor:innen“, „Technische Durchführbarkeit“ der Arbeit mit der Plattform, „Zielgruppenadäquatheit“, „Durchführungsmodus und Schwierigkeit der Aufgaben“, „Teilnahmebereitschaft der Kinder“ und „Feedback aus den Schulen“, mit denen die Tutor:innen kooperieren. Hier galt es, ein möglichst breites Spektrum anschaulicher Erfahrungen und Einschätzungen zu erfassen, so dass standardisierte Antwortvorgaben extensiv um offene Antwortformate für die Spezifizierung der Angaben erweitert wurden.²

Kooperation mit TFA und technische Durchführbarkeit der Plattform: In diesem Themenbereich geht es um Bedingungen, die im Wesentlichen in der Verantwortung

¹ Die ANOVA ergab bei einem F -Wert von 4,627 ($df=2$) mit $p=0,017$ signifikante Mittelwertsunterschiede für die Wirksamkeits-Skala zwischen den drei Gruppen ($n=37$) mit Erfahrung in der Leseförderung (1), „[...] innerhalb der letzten 3 Monate [...] erste Tätigkeit im Rahmen der Leseförderung“, (2), „[...] zwischen drei Monaten und einem Jahr [...]“ sowie (3), „[...] seit mehr als einem Jahr [...]“. Die Levene-Statistik stützt die Annahme der Varianzhomogenität ($p=0,637$). Ein signifikanter Unterschied ist nach dem Bonferroni Post-Hoc-Test zwischen den Gruppen (1) und (3) anzunehmen ($p=0,043$).

² Antworten auf offene Fragen sind im Anhang II, S. 44ff. dokumentiert.

von TFA liegen. Sie sollen Stärken und Probleme in der Zusammenarbeit mit den Tutor:innen und ggf. Hinweise auf Verbesserungsmöglichkeiten aufzeigen.

Alle Tutor:innen werden von TFA in die Grundlagen und die Anwendung der Plattform und des Tutorings eingeführt. Wir haben die Befragungsteilnehmenden gefragt, ob die Einführungsschulung in „Lesen mit dem Turbo-Team“ ausreichend auf den Einsatz vorbereitet hat (Abbildung 4). Weniger als die Hälfte (16 von 38 Personen) bejahte dies uneingeschränkt und hatte den Eindruck, das Tutoring anschließend selbständig durchführen zu können. Fast ein Drittel der Befragten (11 von 38) hat sich erst im Verlauf der praktischen Durchführung des Tutorings zurechtgefunden, und für sieben Tutor:innen stellten sich nach der Schulung noch immer Fragen zur technischen Bedienung der Plattform. In Einzelfällen gab es nach der Schulung noch Fragen zur Verwaltung der Teams sowie zum Inhalt der Förderung und zum Ablauf des Tutorings. Dass sich mehr als die Hälfte der Befragten nach der Schulung noch nicht in der Lage sah, das Tutoring ohne weitere Informationen selbständig durchzuführen, ist ein deutlicher Hinweis darauf, die Gestaltung der Einführung in „Lesen mit dem Turbo-Team“ zu optimieren. Insbesondere ist es ratsam, die Tutor:innen über einen Zeitraum im Prozess zu begleiten, da sich zusätzliche Fragen erst in der Praxis der Anwendung stellen.

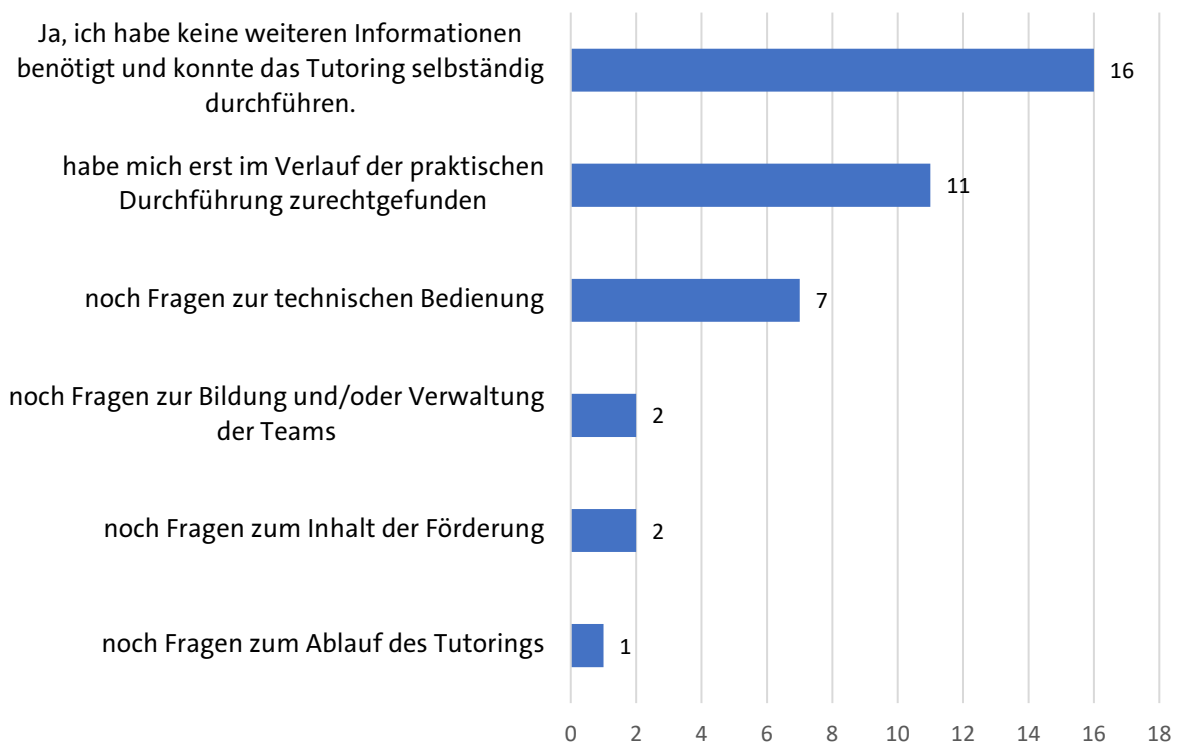


Abbildung 4: Hat Sie die Einführungsschulung in „Lesen mit dem Turbo-Team“ ausreichend auf den Einsatz vorbereitet? (n=38, Mehrfachnennungen bei Antworten 2ff. möglich)

Auch in ihren offenen Antworten zeigen sich die Befragten wohlwollend, freundlich und konstruktiv. Offenbar liegt ihnen daran, einen Beitrag zur Optimierung des gesamten Angebots zu leisten – auch mit Blick auf die Voraussetzungen für eine

erfolgreiche Durchführung, die auf Seiten des dafür eingesetzten Personals gegeben sein müssen.

Danach gefragt, in welcher Hinsicht die Einführungsschulung verbessert werden könne, wurde von einigen Teilnehmenden die Abstraktheit der Informationen angesprochen. Diese wurde zwar nicht rundweg negativ bewertet, aber es scheint für die Überarbeitung der Schulung ratsam, abstrakte Informationen und konkrete Beispiele sowie Übungen mit der Plattform noch besser zu verbinden. In der Bemerkung einer Teilnehmerin wird dies diskret zum Ausdruck gebracht:

„Die theoretische Vorbereitung war sehr informativ, allerdings waren die ersten Schritte in der Praxis turbulent [...].“

Dieser Hinweis spricht auch dafür, die Schulung nicht auf einen einzelnen Termin zu beschränken, sondern Schritte der Einführung und Übung aufeinander folgen zu lassen.

Insgesamt scheint die Relation zwischen Informationen, die bei der Schulung angeboten wurden, und handlungsvorbereitenden Praktiken noch der Optimierung zu bedürfen. Kritisiert wird einerseits, dass die Menge der gebotenen Informationen zu gering sei (*„Etwas kürzen, es waren relativ wenige Informationen über einen zu langen Zeitraum“*). Zugleich aber wird darauf hingewiesen, dass *„Rückfallmöglichkeiten“* benötigt werden, um sich der Anforderungen und Möglichkeiten des Tutorings nach der Schulung weiter vergewissern zu können:

„Für mich war das Tutoring anschaulich und interessant, hat aber für mich für den transportierten Inhalt zu viel Zeit eingenommen. Mir haben die beigelegten Materialien geholfen, weil ich mir nicht alle Inhalte aus der Einführungsschulung nicht merken konnte.“

Dies könnte als Widerspruch gelesen werden, deutet aber aus unserer Sicht eher auf den angesprochenen Aspekt der notwendigen Begleitung der Tutor:innen in ihrer Praxis. Hier zeigt sich das in Qualifizierungsmaßnahmen für pädagogische Berufe (und nicht nur da) immer wieder auftauchende Problem, dass die Wirksamkeit einmaliger Angebote begrenzt ist und nicht genügend Nachhaltigkeit erwarten lässt (Grossman & Salas 2011; Lipowski 2004).

Zu diesem Problem finden sich in den Antworten der Befragten auch Lösungsvorschläge, zum Beispiel:

„Es wäre hilfreich in Präsenz an den Tablets/PCs richtige Einheiten durchzuführen. Man könnte z.B. einzelne TutorInnen abstellen und online schulen, die dann in Präsenz die unsicheren Personen im Team an die Arbeit heranzuführen.“

Angeraten wird hier also ein Multiplikatorenmodell, auf dessen Basis begleitende Angebote von erfahrenen Personen für Novizen gemacht werden. Solche Modelle haben sich in einigen Kontexten bewährt, allerdings auch nur unter Einhaltung von

Bedingungen – etwa: dass die „erfahrenen Personen“ ebenfalls für die Aufgabe geschult werden.

Begleitende Unterstützung durch TFA: Die Frage „Wie erleben Sie die begleitende Unterstützung durch TFA?“ haben 33 Befragte beantwortet. Sie konnten in Analogie zu Schulnoten sechs Beurteilungen zwischen „sehr gut“ und „ungenügend“ abgeben. Insgesamt fiel das Urteil sehr positiv aus: Drei Viertel der Antwortenden gaben hier „gut“ oder „sehr gut“ an, die Noten „mangelhaft“ oder „ungenügend“ wurden nicht vergeben (Abbildung 5).

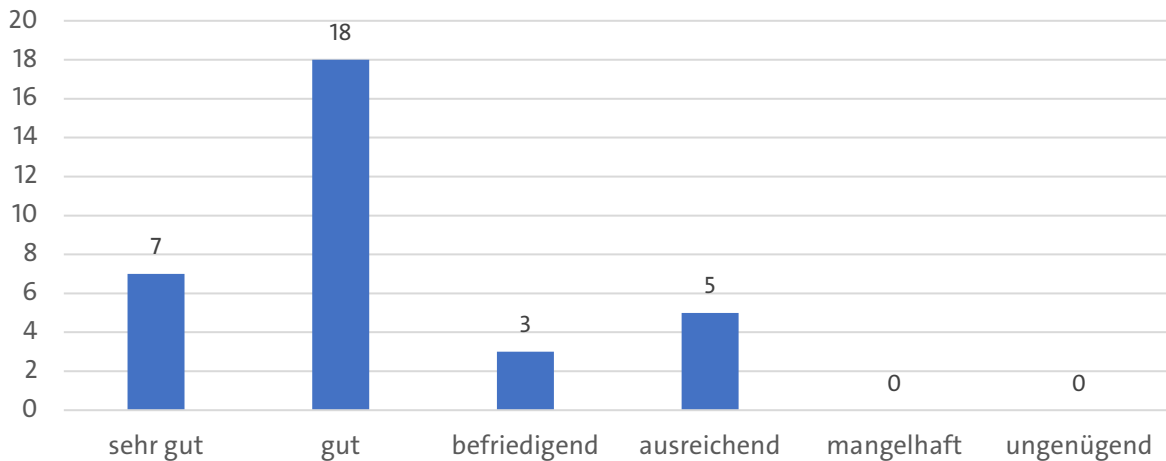


Abbildung 5: Wie erleben Sie die begleitende Unterstützung durch TFA (n=33)

Die Befragten sind also auch hier wohlwollend und an der Optimierung einer schon als gut erinnerten Erfahrung interessiert. Die Nachfrage „In welcher Hinsicht könnte die Unterstützung durch TFA verbessert werden?“ konnte wieder in einem offenen Format beantwortet werden. Mehrfach gewünscht wird die „Einführung von festen Telefonsprechzeiten“, aber auch das Interesse an einer Teambildung der Gruppe der Tutor:innen wird geäußert: „Regelmäßiger Austausch mit ALLEN im Team, nicht nur mit den leitenden Personen.“

Technische Durchführbarkeit: Die Frage „Wie schätzen Sie die technische Durchführbarkeit der Spiele und Aufgaben und die Stabilität der Internetverbindung ein?“ konnte mit einer der drei Antwortmöglichkeiten „Durchweg oder überwiegend gut.“, „Teilweise gut, bei einigen Spielen oder zeitweise weniger gut.“ oder „Überwiegend oder durchweg schlecht.“ beantwortet werden. Letztere Antwortmöglichkeit wurde nicht genannt. Deutlich mehr als die Hälfte (19 von 32) schätzten die technischen Voraussetzungen als „durchweg oder überwiegend gut“ ein, allerdings bewerteten 13 Tutor:innen die technische Durchführbarkeit als „teilweise gut, bei einigen Spielen oder zeitweise weniger gut“.

Die offene Folgefrage „Wo genau gibt es Probleme mit der technischen Durchführbarkeit der Spiele und Aufgaben?“ bringt zutage, dass technische

Schwierigkeiten vornehmlich auf die unzureichende technische Ausstattung an vielen Schulen zurückzuführen war. Anwendungsproblemen gab es insbesondere, weil das WLAN instabil war und dadurch Verbindungsprobleme auftraten:

„In manchen Schulen gibt es einen schlechten Internetempfang. In solchen Fällen dauert es manchmal länger oder gibt es Probleme die Spiele zu spielen. Vor allem bei dem Spiel "Wörter Suche", weil da ein Video abgespielt wird.“

Auch die Ausstattung mit digitalen Geräten scheint noch eine Hürde zu bereiten, wie hier geschildert wird:

„Die verwendeten Tablets der Schule sind nicht auf einem einheitlichen Stand (Einstellungen) und sind dadurch nicht jederzeit unmittelbar einsatzbereit (z.B. kein Ton, Bildausschnitt passt nicht in den Bildschirm. Die Anwendung hängt sich auf bzw. Verbindungsprobleme mit dem Internet. Nicht alle Räume haben Internetzugang. Verzögerung zwischen Eingabe und Aktionsausführung - dies führt gerade im Spiel "Schreibstar" zu "unechten" Fehlern und Frustration bei den Schülern.“

Die technische Ausstattung von Schulen ist – so steht zu hoffen – ein Problem, das sich in absehbarer Zeit beheben lassen wird. In der Zwischenzeit wäre es aber ratsam, bei Vereinbarungen über die Anwendung von TFA vorab eingehender zu klären, ob die technischen Voraussetzungen dafür vorhanden sind.

Zielgruppenpassung: Bei den Fragen zu diesem Themenfeld geht es um Einschätzungen der Tutor:innen zum Problem der Adaptivität von TFA an die Kinder, die in den Genuss des Trainings kommen sollen. Relevant aus unserer Sicht sind hier Fragen nach der Passung von Spielen und Aufgaben zum Alter der Kinder und dazu, durch welche Teile des Trainings die Kinder ggf. über- oder auch unterfordert sind. Zudem ist es für die Einschätzung der Tauglichkeit des dialogischen Prinzips des Tutoring relevant, ob die Kinder in der Praxis des Tutorings besser im Team oder einzeln arbeiten.

Die Frage „Sind die Spiele und Aufgaben für die von Ihnen geförderten Kinder altersgerecht?“ beantworten 14 von 32 Befragten mit „durchweg oder überwiegend ja“, 16 mit „teilweise ja, andere weniger“ und zwei mit „überwiegend oder durchweg nein“ (Tabelle 8). Aus der Tabelle geht auch hervor, dass eine deutliche Mehrheit derjenigen Tutor:innen, die „Lesen mit dem Turbo-Team“ auch mit Schüler:innen ab Klasse 6 durchführen, die Altersgemäßheit zumindest von Teilen der Spiele und Aufgaben anzweifeln – was zu erwarten war. Gut die Hälfte derjenigen, die – entsprechend der bei der Pilotevaluation vorliegenden Fassung des Tutorings – nur mit jüngeren Kindern arbeiten, schätzt die Aufgaben für durchweg oder überwiegend altersgerecht ein, aber auch unter ihnen finden sich zehn Personen, die Bestandteile des Programms, als nicht altersgemäß empfinden.

Tabelle 8

Sind die Spiele und Aufgaben für die von Ihnen geförderten Kinder altersgerecht? Antworten nach Tutor:innen mit Schüler:innen in der Primarstufe, Primar- und Sekundarstufe und Sekundarstufe

	nur Klasse 1 bis 5	(auch) Klassen 6 bis 8	Gesamt
Durchweg oder überwiegend ja.	12	2	14
Teilweise ja, einige weniger.	9	7	16
Überwiegend oder durchweg nein.	1	1	2
Gesamt	22	10	32

Nicht überraschend ist, dass sich die Antworten auf offene Fragen zur Passung zum Alter der Lernenden überwiegend auf die geringe Eignung für ältere Schüler:innen beziehen. Hier gibt es ebenso Anmerkungen zur optischen Gestaltung, die für ältere Kinder nicht ansprechend sei, als auch zur sprachlichen Form der Präsentation von Aufgaben, die sich eher an Grundschulkindern richte, und schließlich Hinweise auf Interessengebiete älterer Schüler:innen, die nicht berücksichtigt seien.

„Das ganze Programm ist weder inhaltlich noch optisch altersgemäß für Kinder, die älter als 12 sind. Dafür ist es zu kindlich gestaltet. Kinder der Klassenstufe 7 oder älter werden von dem Programm nicht angesprochen.“

Eine alternative Einschätzung ist zwar auch vertreten:

„Die SuS mögen spielen, egal wie alt sie sind. Ich sehe dabei überhaupt kein Problem. Außerdem ist der Wortschatz im Kindergartenalter genauso wichtig wie der im Grundschulalter, usw. Mit diesem Programm haben die DaZ-SuS Möglichkeit, unterschiedliche Wortschatzebenen kennen zu lernen, anzueignen.“

Insgesamt aber sind die Rückmeldungen zu diesem Themenbereich eher für die Weiterentwicklung von TFA geeignet als für die Verbesserung der vorliegenden Fassung für die Grundschule. Indirekt immerhin, wird die Fassung für das Grundschulalter positiv gewürdigt:

„Oft ist es eher an den Bedürfnissen der Grundschulkindern gerichtet, so dass sich die Schüler der weiterführenden Schule nicht angesprochen fühlen.“

Über- und Unterforderung: Tabelle 9 fasst die Ergebnisse zu den beiden Fragen zu diesem Aspekt zusammen. Bei Antwort „ja“ wurde jeweils nachgefragt, welche Spiele die Kinder regelmäßig über- bzw. unterfordern.

Tabelle 9

Gibt es Spiele, die die von Ihnen geförderten Kinder regelmäßig überfordern bzw. unterfordern?

	Überforderung	Unterforderung
Bisher nicht aufgetreten	28	18
Ja [...]	4	14
Gesamt	32	32

Eine regelmäßige Überforderung wird nur in vier Fällen genannt. Die offenen Nennungen lassen keine systematischen Tendenzen erkennen.

Eine regelmäßige Unterforderung wird dagegen von 14 Tutor:innen wahrgenommen. Dabei sind diejenigen, die Kinder höherer Klassenstufen fördern, keineswegs in der Mehrheit, sodass die Antworten nicht deckungsgleich mit denen zur Altersgemäßheit sind. Mehrfach werden die Spiele „ABC-Profi“ und „ABC-Sprint“ als unterfordernd wahrgenommen, wobei auch die Gefahr des Verlusts an Interesse gesehen wird:

„Das Buchstabieren (Aussprechen einzelner Buchstabenkombinationen) wird teilweise als sehr langweilig empfunden.“

Hingewiesen wird auch auf mögliche Unterforderung durch die Kombination des verlangten Typs der Aktivitäten mit dem inhaltlichen Interesse der Schülerinnen und Schüler:

„Leseblitz: die SuS lesen teilweise nicht mit, schalten ab. Ich vermute, dass es am (altersabhängigen) Interesse am Thema und dem fehlenden spielerischen Charakter oder dem Lesen selbst (und die Wiederholung des Spieles) liegt.“ (nur Kinder der Klassenstufen 1 bis 5)

Tutoring im Team: Ein wesentlicher Aspekt bei der Gestaltung des Tutorings ist die Anlehnung an Prinzipien des kooperativen Lernens, die zu den Grundlagen für das in den USA entwickelte Vorbild von TFA gehören (Slavin 1995). Unterstützt wird die Relevanz dieses Ansatzes durch Forschung zur Sprachentwicklung und Sprachförderung, die die Bedeutung dialogischer Unterstützung des Prozesses hervorhebt (Tracy 2023). Daher ist das Tutoring so aufgebaut, dass prinzipiell jeweils zwei Kinder als Teamspieler beteiligt sein sollen, die sich auch gegenseitig „coachen“ können. Im Ausnahmefall, etwa wenn temporär Kinder im Krankheitsfall ausfallen, kann das Tutoring auch im „Single-Modus“ stattfinden. In der Praxis der Befragten scheint der Ausnahmefall durchaus nicht selten vorzukommen, da die meisten befragten Tutor:innen eine Einschätzung dazu hatten, ob die Durchführung des Tutorings mit Teams oder mit Einzelspieler:innen besser funktioniert.

Nur vier von 32 Tutor:innen waren der Meinung, dass das Tutoring im Team besser funktioniert. Demgegenüber zogen acht Personen das Tutoring mit Einzelspieler:innen vor. Die große Mehrheit von 20 Befragten gab an, dass das Funktionieren des Tutorings

von der Zusammenstellung der Teams abhängt. Diese Beobachtung ist vielfach durch Forschung belegt (Decristan et al. 2021). Die Einschätzungen der Tutor:innen sind insofern nicht überraschend. Als Probleme der Durchführung im Team werden die Zusammensetzung aus Kindern unterschiedlicher Leistung, gegenseitige Ablenkung oder Streit unter den Teammitgliedern genannt. Als Problemlösungen nannten die Befragten, dass sie selbst als Teampartner fungieren oder dass die Kinder doch einzeln arbeiten.

„Das gegenseitige coachen funktioniert nicht immer, besonders wenn die Niveaus nicht komplett identisch sind. Teilweise war es besser, wenn ich als Betreuer eingesprungen bin.“

„Die Gruppen sind teilweise sehr heterogen. Daher macht das Spielen im Single-Modus mehr Sinn. Selten sind Kinder genau gleich eingestuft und es ist nicht ganz fair, wenn einige Kinder in Teams spielen dürfen und andere alleine vor den Tablets sitzen. Die Einteilung ist eine große Herausforderung in der Praxis.“

Aus den Antworten spricht auch, dass es offenbar nicht zu den vertrauten Praktiken für die Kinder gehört, in Teams zu kooperieren.

„Unterschiedliche Leistungsstufen und Frustrationsgrenzen. Gegenseitige "Sabotage", Übergenaugigkeit bzw. "das Suchen von Fehlern" demotivieren die Spieler und erhöht die Frustration.“

Die Antworten deuten darauf, dass die mit kooperativen Ansätzen verbundenen Anforderungen sowohl den Tutor:innen als auch den Kindern wenig vertraut sind. Für beide Gruppen ist das Vorgehen voraussetzungsreich: Kinder müssen „Regeln“ für Formen der Kooperation kennengelernt haben, damit sie in Teams konstruktiv agieren. Tutor:innen müssen die Kenntnisse und die Fähigkeit besitzen, Tandems sinnvoll zusammenzusetzen und mit den Regeln guter Zusammenarbeit vertraut zu machen, sofern die Beteiligten entsprechende Umgangsformen nicht bereits aus ihrer alltäglichen Unterrichtsumgebung mitbringen. Die mit der Zusammenstellung von Teams und der Arbeit mit ihnen verbundenen Herausforderungen sind offenbar ein Thema, das in der Schulung und der Betreuung der Tutor:innen durch TFA ein größeres Gewicht erhalten sollte. Dabei geht es sowohl um Informationen über kooperatives Arbeiten als Prinzip des Tutorings als auch um die Unterstützung bei praktischen Fähigkeiten zur Zusammensetzung der Tandems und zur Begleitung der Kinder auf dem Weg zur Fähigkeit, konstruktiv zusammenzuarbeiten.

Einschätzung der Teilnahmebereitschaft der Kinder: Zu den Ansprüchen von TFA gehört, dass die Kinder gern mit dem Angebot arbeiten und zur regelmäßigen Teilnahme motiviert sind. Daher wurde den Tutor:innen die Frage gestellt: „Wie schätzen Sie die Teilnahmebereitschaft der Kinder für ‚Lesen mit dem Turbo-Team‘ im Allgemeinen ein?“. Die Antworten zeigen, dass gerade zu Beginn der Förderung eine hohe

Teilnahmebereitschaft der Kinder besteht, auch wenn sie im weiteren Verlauf oft nicht in gleicher Höhe aufrechterhalten werden kann. Ein Abflachen der Motivation mit der Zeit ist allerdings erwartbar, wenn sich der Reiz des Neuen legt. Zumindest kann davon ausgegangen werden, dass das Programm das Interesse der Kinder an dieser Förderung weckt. Dass die Motivation später nicht ganz verloren geht, zeigen die zur „Wirksamkeitseinschätzung“ angeführten Antworten auf das Item „‘Lesen mit dem Turbo-Team‘ wirkt motivierend auf meine Schüler(innen)“, das eine relativ hohe mittlere Zustimmung der Tutor:innen erfahren hat.

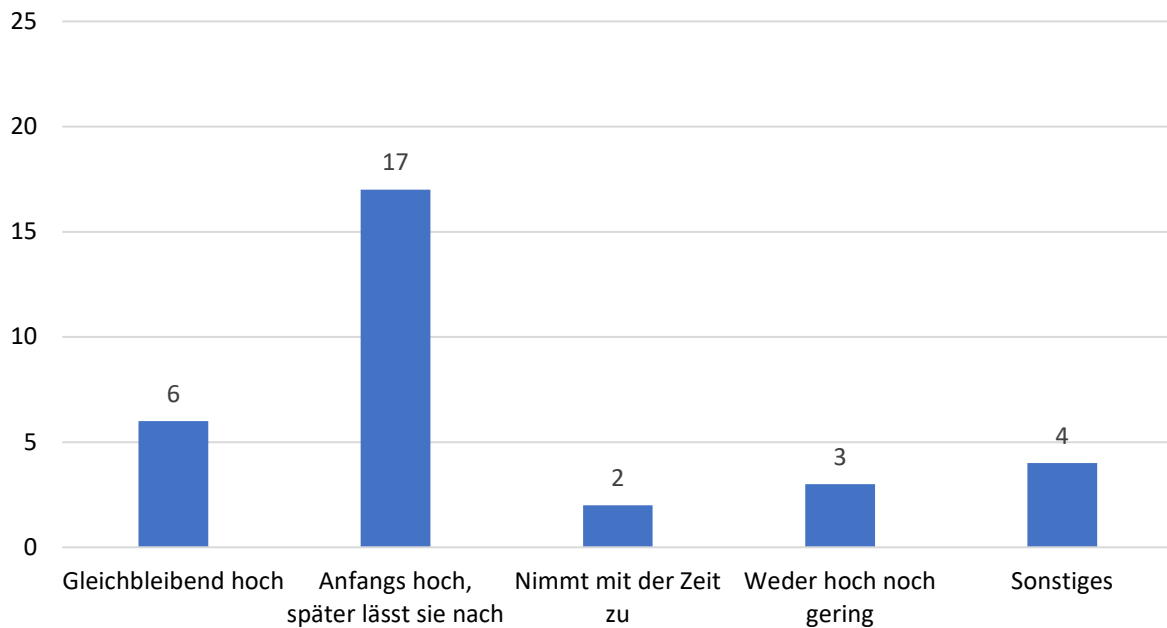


Abbildung 6: Wie schätzen Sie die Teilnahmebereitschaft der Kinder für „Lesen mit dem Turbo-Team“ im Allgemeinen ein? ¹ (n=32)

6.3. Diskussion der Befragungsergebnisse

Aus der Online-Befragung der Tutor:innen ergeben sich deutliche Hinweise auf die Akzeptanz des Programms durch sie selbst und die teilnehmenden Kinder, zur Einschätzung von Lernerfolgen und zu weiteren Erfahrungen, die in der Praxis gemacht werden, inklusive solcher, die als problematisch wahrgenommen werden. Bei der Bewertung der Ergebnisse ist zu berücksichtigen, dass etliche Befragte noch nicht sehr lange mit dem TFA-Programm gearbeitet haben. Dennoch sollten die vorliegenden Hinweise wertvoll für die Weiterentwicklung der Plattform sowie für die künftige Ausgestaltung der Schulung und Begleitung von Tutor:innen und schließlich dafür sein,

¹ Die vorgegebene Antwortmöglichkeit „eher gering“ wurde nicht angekreuzt und ist in der Abbildung nicht dargestellt. Die unter „Sonstiges“ ergänzten Antworten weisen auf unterschiedliche und individuelle Verläufe hin.

das zugrundeliegende Prinzip des kooperativen Lernens in der Praxis des Tutorings umzusetzen.

Die hohe Akzeptanz, die „Lesen mit dem Turbo-Team“ unter den Tutor:innen erfährt, ist eine solide Basis für eine erfolgreiche Realisierung der Förderung. Dieser Eindruck wird gestützt durch die positive Einschätzung der Wirksamkeit von TFA. Die Wertschätzung des Programms zeigt sich besonders bei eher erfahrenen Förderkräften, die bereits mit anderen Förderansätzen gearbeitet haben. Neueinsteiger:innen äußern sich erwartungsgemäß vorsichtiger. Die Anwendung des TFA-Programms zum Zeitpunkt der Befragung fand also in einem Umfeld statt, das ihm grundsätzlich wohlgesonnen ist. Dies spricht auch aus der allgemein sehr positiven Bewertung der begleitenden Unterstützung durch das TFA-Team. Deutlich wird in den Antworten das Interesse der Befragten an einer weiteren Optimierung des Tutorings und der Möglichkeit, daran mitzuwirken.

Damit ist auch aus dieser Sicht mit der Plattform in der Version, die in die Pilotevaluation einging, ein sehr guter Anfang für die Entwicklung einer akzeptierten und wirksamen Maßnahme gemacht, die Schülerinnen und Schüler zum erfolgreichen Lesenlernen verhelfen kann, die – aus welchen Gründen auch immer – Hürden im Lernprozess zuvor nicht überwinden konnten.

Die vorgestellten Ergebnisse betreffen ausschließlich die Fassung der Tutoring-Plattform, die zum Zeitpunkt der Durchführung der Pilotevaluation vorlagen. Inzwischen sind Weiterentwicklungen vorgenommen worden, etwa zum Einsatz der Plattform für Tutoring am Anfang der Sekundarstufe (<https://tutoringforall.de/>). Zur Einschätzung von Weiterungen, die inzwischen vorgenommen wurden – etwa mit Blick auf den Einsatz bei Schülerinnen und Schülern im Sekundarbereich – besitzen wir auf der Basis unserer Pilotevaluation keine Informationen. Das Entwicklungsteam sollte den einmal eingeschlagenen Weg weiterhin konsequent zu beschreiten, dass sorgsame und wissenschaftlichen Standards (so gut, wie im jeweiligen Augenblick möglich) einhaltende Evaluationen zur Klärung der Frage der Wirksamkeit und zur prozessbegleitenden Verbesserung der Qualität der Maßnahme realisiert werden.

Literatur

- Bortz, J. & Döring, N. (2002): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*: Heidelberg: Springer
- Chambers, B., Abrami, P., Tucker, B., Slavin, R., Madden, N., Cheung, A., & Gifford, R. (2008). Computer assisted tutoring in Success for All: Reading outcomes for first grade. *Journal of Research on Effective Education*, 1 (2), 120-137.
- Decristan, J., Schastak, M., Reitenbach, V. & Rauch D. (2021). Implementation von Peer Tutoring unter Berücksichtigung von Mehrsprachigkeit. *Zeitschrift für Erziehungswissenschaft*, 24, 841-860.
- Fahrmeir, L., Kneib, T.; Lang, S. (2009). *Regression: Modelle, Methoden und Anwendungen*. 2. Aufl., Berlin: Springer.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Ganzeboom, H. B.G. (2010). *A New International Socio-Economic Index (ISEI) of occupational status for the International Standard Classification of Occupation 2008 (ISCO-08) constructed with data from the ISSP 2002–2007*. Paper presented at Annual Conference of International Social Survey Programme, Lisbon, May 1, 2010.
https://www.researchgate.net/publication/242660186_A_new_International_Socio-Economic_Index_ISEI_of_occupational_status_for_the_International_Standard_Classification_of_Occupation_2008_ISCO-08_constructed_with_data_from_the_ISSP_2002-2007 [accessed Jul 05 2024]
- Ganzeboom, H. B.G., & Treiman, D. J. (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. In J. H.P. Hoffmeyer-Zlotnik & C. Wolf (Eds.): *Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables*. New York: Kluwer Academic / Plenum Publishers, 159-193.
- Grossman, R. & Salas, E. (2011). The transfer of training: What really matters. *International Journal of Training and Development*, 15(2), 103-120.
- Haack, G. & Haß, W. (2024). Evaluation. In: Bundeszentrale für gesundheitliche Aufklärung (BZgA) (Hg.). *Leitbegriffe der Gesundheitsförderung und Prävention. Glossar zu Konzepten, Strategien und Methoden*.
<https://doi.org/10.17623/BZGA:Q4-i016-3.0>
- Hornberg, S., Naltin, R., Potthoff, B., Schwippert, K., Schulz-Zander, R. (2007). Lesekompetenzen von Mädchen und Jungen im internationalen Vergleich. In Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., Schwippert, K. & Naltin, R. (Hg.): *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann, 195-223.

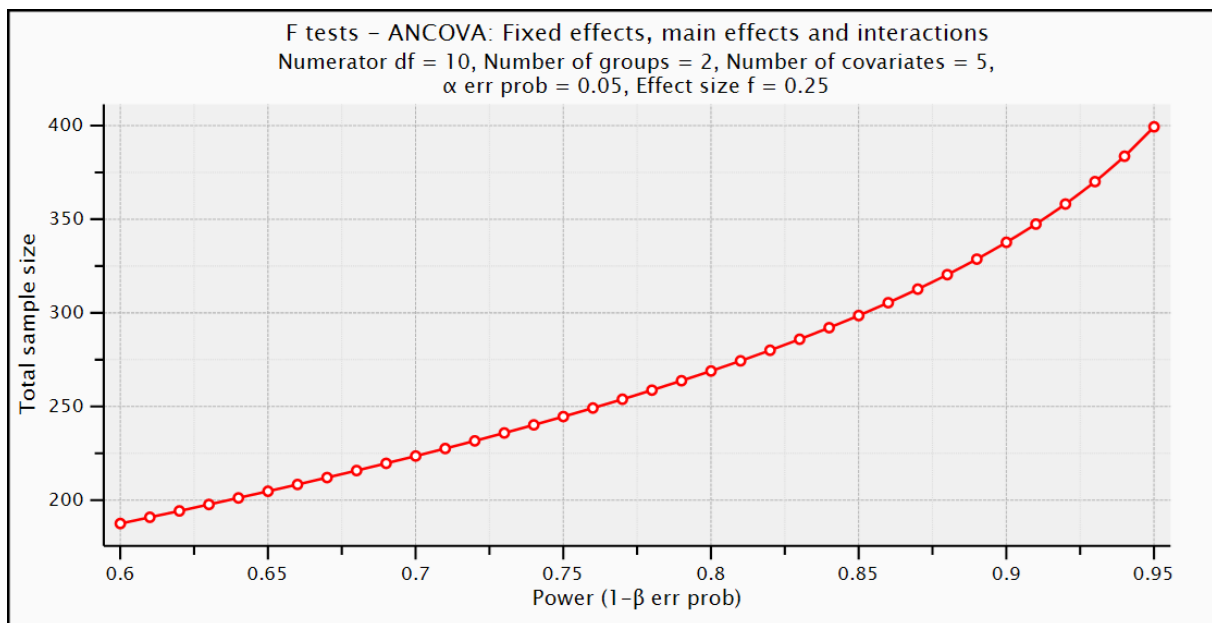
- IQTIG – Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (Hg.) (2021). Konzept für eine onlinebasierte Patientenbefragung im Rahmen der sektorenübergreifenden Qualitätssicherung nach § 136 SGB V. Abschlussbericht. Berlin. https://iqtig.org/downloads/berichte/2021/IQTIG_Konzept-onlinebasierte-Patientenbefragung_Abschlussbericht_2021-11-19.pdf
- Little, R.J.A. & Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.
- Lipowsky, F. (2004). Was macht Fortbildungen für Lehrkräfte erfolgreich? Befunde der Forschung und mögliche Konsequenzen für die Praxis. *Die Deutsche Schule*, 96(4), S. 462-479.
- Madden, N. A., & Slavin, R. E. (2017). Evaluations of Technology-Assisted Small-Group Tutoring for Struggling Readers. *Reading & Writing Quarterly*, 1-8. <http://dx.doi.org/10.1080/10573569.2016.1255577>.
- Mayerl, J. & Urban, D (2019). Vorsicht (!) bei Regressionsanalysen mit Interaktionsvariablen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 71: 135–156.
- Mayringer, H. & Wimmer, H. (2014). *SLS2-9–Salzburger Lesescreening für die Schulstufen 2-9*. Göttingen: Hogrefe.
- Meudt, S.-I., Zeuch, N., Neuber, L. & Souvignier, E. (2020). Kurzskalen zur Erfassung des Transfererfolgs von Leseförderkonzepten in der Schulpraxis. In Mackowiak, K., Beckerle, C., Gentrup, S. & Titz, C. (Hg.): *Forschungsinstrumente im Kontext institutioneller (schrift-)sprachlicher Bildung*. Bad Heilbrunn: Julius Klinkhardt, 141-157. URN: urn:nbn:de:0111-pedocs-201480 - DOI: 10.25656/01:20148
- Möbus, C. & Nagl, W. (1983). Messung, Analyse und Prognose von Veränderungen. In Bredenkamp, J. & Feger, H. (Hg.). *Hypothesenprüfung*. Göttingen: Hogrefe. DOI: 10.13140/RG.2.1.1303.7283
- Rosebrock, C. & Nix, D. (2020): *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung*. Baltmannsweiler: Schneider Hohengehren.
- Ross, S. M., Laurenzano, M. & Madden, N.A. (2017). *An Evaluation of the Lightning Squad Computer-Assisted Small Group Tutoring Program on the Reading Achievement of Disadvantaged Students in Grades 1-3*. Technical Report. Center for Research and Reform in Education, Johns Hopkins University.
- Rossi, P.; Freeman, H.; Hofman, G. (1988): *Programm-Evaluation: Einführung in die Methoden angewandter Sozialforschung*. Stuttgart: Enke.
- Slavin, R. (1995). *Cooperative learning. Theory, research, and practice*. Boston u.a.: Allyn and Bacon.
- Theobald, A. (2017). *Praxis Online-Marktforschung. Grundlagen - Anwendungsbereiche - Durchführung*. Wiesbaden: Springer Gabler.

Tracy, R. (2023). Erwerb sprachlicher Kompetenzen im Lebensverlauf. In M. Becker-Mrotzek, I. Gogolin, H-J. Roth und P. Stanat (Hg.): *Grundlagen der sprachlichen Bildung*. Münster: Waxmann, S. 57–83.

Urban, D., Mayerl, J., & Wahl, A. (2016). *Regressionsanalyse bei fehlenden Variablenwerten (missing values): Imputation oder Nicht-Imputation? Eine Anleitung für die Regressionspraxis mit SPSS*. Arbeitspapier. Stuttgart: Universität Stuttgart, Institut für Sozialwissenschaften, Abt. f. Soziologie u. empirische Sozialforschung. <http://dx.doi.org/10.18419/opus-9038>.

Anhang I: Power-Analyse mit G*Power

Das von uns präferierte Regressionsmodell entspricht einem Kovarianzmodell, das den Effekt einer Maßnahme als Unterschied zwischen zwei Gruppen unter Kontrolle von maximal fünf Kovariaten misst. Dabei nehmen wir konventionell einen α -Fehler von 5 Prozent (=Signifikanzniveau) und eine Power oder Teststärke von 80 Prozent (= β -Fehler von 20 Prozent) an¹. Gehen wir von einer mittleren Effektstärke von $f=0,25$ aus, dann lässt sich in der unten gezeigten und mit G*Power (Faul et al. 2007) erstellten Grafik für eine Power von 80 Prozent eine Stichprobengröße von 269 Personen ablesen.



Grafik erstellt mit G*Power ©

¹ α -Fehler: Wahrscheinlichkeit, die Nullhypothese abzulehnen, obwohl sie richtig ist.
 β -Fehler: Wahrscheinlichkeit, die Nullhypothese beizubehalten, obwohl sie falsch ist.

Anhang II: Antworten auf offene Fragen in der Tutor:innenbefragung

a) *In welcher Hinsicht könnte Ihrer Ansicht nach die Einführungsschulung verbessert werden?*

Die Materialien haben mir beim Hineinfinden geholfen, könnte aber kompakter sein (Quickstart-Guide?). Die Einführungsschulung hat mir geholfen, den Ablauf des Tutorings kennenzulernen. Es wäre praktisch, würde eine Videoaufzeichnung zur Verfügung gestellt werden, wenn man einen bestimmten Schritt in Vorbereitung noch einmal ansehen möchte. Außerdem könnte man überlegen, die Teilnehmer in "Anfänger" und "Fortgeschrittene" zu unterteilen. Für mich war das Tutoring anschaulich und interessant, hat aber für mich für den transportierten Inhalt zu viel Zeit eingenommen. Mir haben die beigelegten Materialien geholfen, weil ich mir nicht alle Inhalte aus der Einführungsschulung nicht merken konnte. Vielleicht ist das aber auch ein "ich-Problem".

Die theoretische Vorbereitung war sehr informativ, allerdings waren die ersten Schritte in der Praxis turbulent (Testen der SuS, Tiernamen merken, Teams bilden...alles musste eingespielt werden.

Die Unterstützung durch [REDACTED] war hervorragend

DIE war in Ordnung.

Einfachere Handhabung bei dem Anlegen der Gruppen

Es wäre hilfreich in Präsenz an den Tablets/PC's richtige Einheiten durchzuführen. Man könnte z.B. einzelne TutorInnen abstellen und online schulen, die dann in Präsenz die unsicheren Personen im Team an die Arbeit heranführen.

Es wurde sich genügend Zeit genommen.

Etwas kürzen, es waren relativ wenige Informationen über einen zu langen Zeitraum.

Für mich war diese Umfrage schwer, weil ich mit Kindern aus der VKL arbeite, das heißt sie sprechen kaum Deutsch.

In dem klar verdeutlicht wird, ob und wie das Programm für die Klasse 1 überhaupt geeignet ist.

Indem vor dem Training eine Vorstellung vom Programm gegeben werden und die Leitfaden zum Vorlesen eingesendet werden.

Inhalt der Berichte; Hinweise zur Handhabung bei Wechsel des Tutors, Zusammenhänge zwischen jeweiliger Lesestufe/Lesefähigkeit und Ziel-Wert für RWM

ja, aber teilweise langwierig

Mehr Informationen zu den einzelnen Berichten.

Vielleicht könnten Illustrationen zu den unbekanntem Wörtern hinzugefügt werden.

wie man etwas löscht, dass versehentlich falsch war

Wir haben im Kleinteam nochmals die Funktions- und Arbeitsweise durchgestrichen, dadurch waren dann die letzten Unklarheiten beseitigt. Es war schon teils recht abstrakt, im Endeffekt war es im ersten Durchgang dann doch selbsterklärend

b) In welcher Hinsicht könnte die Unterstützung durch TFA verbessert werden?

Einführung von festen Telefonsprechzeiten

Einloggen auf jedem einzelnen iPad zu Beginn immer sehr zeitaufwendig - Andere Log-In Möglichkeit?

Erfahrungsaustausch könnte eventuell hilfreich sein. Außerdem könnten weitere Schulleitungen mit positiven Ergebnissen überzeugt werden, damit das Projekt bei ihnen auch eingeführt wird.

Feste Telefonsprechzeiten könnten eingeführt werden.

hatte noch keinen Bedarf

In vielen Schulen gibt es kein bzw. schlechtes W-lan so dass diese Kinder dann benachteiligt sind da man das Tutor nicht anwenden kann

komfortablerer Zugang zum Lernprogramm, weitere Materialien zur Unterstützung der Tutoren (Videomitschnitt, Quickstart-Guide)

Regelmäßiger Austausch mit ALLEN im Team, nicht nur mit den leitenden Personen.

Überarbeitung der Aufgaben für Klasse 1, da dies viel zu schwierig und teilweise unlösbar ist
Mehr Auswahl an Geschichten, da den Kindern schon langweilig wird, weil sie die Geschichten schon mehrmals gelesen haben

Würde ich gern weiter ausholen... kann man Ihnen auch eine Mail schreiben? Hab den Fehler gemacht, das hier am Smartphone zu beantworten, und das ist schreibtechnisch nicht günstig (Ich gehöre zur Handy-Einfinger-Tipp-Generation)

c) *Wo genau gibt es Probleme mit der technischen Durchführbarkeit der Spiele und Aufgaben?*

Ab und zu funktioniert das W-Lan nicht

Am iPad wird sehr oft der Vollbildmodus geschlossen - Internetleisten sind zu sehen

Aussprache der Wörter nicht genau bzw. zu umgangssprachlich, wodurch sich Fehler bei den Kindern einschleichen - bei Schreibstar

bei schlechtem W-lan

Bei uns an der Schule gilt die Seite als "nicht vertrauenswürdig" und lädt dementsprechend nicht gut.

Der Ton war hin und wieder weg gelegentlich sind die Spiele abgestürzt.

Der Übergang zwischen den einzelnen Spielen zieht sich etwas. Das könnte gekürzt werden

Die Schulen sind teilweise schlecht ausgestattet mit Wlan. Die App hängt deshalb immer mal wieder oder ist verzogen.

Die verwendeten Tablets der Schule sind nicht auf einem einheitlichen Stand (Einstellungen) und sind dadurch nicht jederzeit unmittelbar einsatzbereit (z.B. kein Ton, Bildausschnitt passt nicht in den Bildschirm. Die Anwendung hängt sich auf bzw. Verbindungsprobleme mit dem Internet. Nicht alle Räume haben Internetzugang. Verzögerung zwischen Eingabe und Aktionsausführung - dies führt gerade im Spiel "Schreibstar" zu "unechten" Fehlern und Frustration bei den Schülern.

Es gab teilweise Probleme beim einloggen, trotz richtiger Passwörter und Zugänge. Das kann einem schon mal die Zeit rauben...

In manchen Schulen gibt es einen schlechten Internetempfang. In solchen Fällen dauert es manchmal länger oder gibt es Probleme die Spiele zu spielen. Vor allem bei dem Spiel "Wörter Suche", weil da ein Video abgespielt wird.

In meiner Schule gibt es keine technischen Probleme: Internet ist stabil und wir haben auch genügend Geräte zur Verfügung gestellt bekommen.

Keine

Kinder können raten

liegt wahrscheinlich eher am Schulinternet

Oft an der Ausstattung der Einrichtungen, in denen es angewendet wird.

Probleme nur wegen des schulinternen W-Lans

Unser Schul Wlan 😊

WLAN in der Schule ist instabil, vereinzelt hängt sich die Seite auf und erfordert erneute Anmeldung, die Arbeit mit Pseudonymen erschwert vor allem bei Teams die Anmeldung der SuS (wegen den Abkürzungen bei der Teamauswahl auf dem Startbildschirm)

d) Welche Spiele und Aufgaben sind Ihrer Ansicht nach nicht altersgerecht?

ABC Profi ist für höhere Stufen manchmal etwas ungeeignet und unterfordernd

ABC-Sprint für 4-Klässler

Buchstaben erkennen/Laute hören Textinhalte... meine Kinder sind Klasse 4 und werden anhand des Textlevels offenbar in Klasse 1 bzw. 2 eingestuft. Das empfinden die Kinder als demotivierend ...

Das ganze Programm ist weder inhaltlich noch optisch altersgemäß für Kinder, die älter als 12 sind. Dafür ist es zu kindlich gestaltet. Kinder der Klassenstufe 7 oder älter werden von dem Programm nicht angesprochen.

Die SuS mögen spielen, egal wie alt sie sind. Ich sehe dabei überhaupt kein Problem. Außerdem ist der Wortschatz im Kindergartenalter genauso wichtig wie der im Grundschulalter, usw. Mit diesem Programm haben die DaZ-SuS die Möglichkeit, unterschiedliche Wortschatzebenen kennen zu lernen, anzueignen.

Durch die Anwendung im DaZ-Bereich sind auch etwas ältere SuS dabei, für die die Sprache zu "kindgerecht" sein könnte

Für ältere Kinder, die schon recht flüssig lesen, wird es evtl. nach einer Weile langweilig

Für ältere Schüler, die Deutsch als Zweitsprache lernen, ist es nicht ganz altersgemäß. Für den Prozess des Lesen-Lernens aber eher zweitrangig. Eventuell ein zweites Geschäftsfeld für diese Zielgruppe eröffnen?!

Für Klasse 1 sind nur ABC Profi geeignet und das auch nur teilweise

Für viele Kinder langweilig, immer das Gleiche zu spielen

Keine

Klasse 6, niedriges Niveau: zu kindlich für bereits pubertierende Kinder/Jgdl.

Oft ist es eher an den Bedürfnissen der Grundschul Kinder gerichtet, so dass sich die Schüler der weiterführenden Schule nicht angesprochen fühlen.

Schlechte Leser:innen bekommen Geschichten, die inhaltlich auf erste oder zweite Klassen zugeschnitten sind.

Unsere Schüler sind meistens zu alt für die Textinhalte

e) Welche Spiele überfordern die Kinder regelmäßig?

Beim gegenseitigen Vorlesen und Markieren der Fehler durch den Coach können die Pärchen gut schummeln, indem sie die fehlerhaft vorgelesenen Wörter nicht markieren.

Die Höhlengeschichte kam bei mehreren Kindern eindeutig zu früh. Sie konnten hier nicht ansatzweise die RWPM erreichen, die sie zuvor bei der Einstufung geschafft haben.

Generell für Kinder Anfang der 1. Klasse, da einige Buchstaben nicht bekannt sind

Wenn die Kinder das Wort nicht kennen

f) Welche Spiele unterfordern die Kinder regelmäßig?

ABC Profi

ABC-Profi

ABC-Sprint

Ältere Kinder langweilen sich manchmal bei langen Texten

Beim gegenseitigen Vorlesen wird oft geschummelt.

Buchstaben lesen - laute aussprechen, und dann immer wieder erst hören (prüfen)müssen, damit auf richtig geklickt werden kann

Das Buchstabieren (Aussprechen einzelner Buchstabenkombinationen) wird teilweise als sehr langweilig empfunden.

Das Spiel mit einzelnen Lauten

Die untersten Level, da die Förderung bei uns in der 5. und 6. Jahrgangsstufe stattfindet.

Es gab Kinder mit der Bodo/Dodo Geschichte, die zu einfach waren.

Ich weiß den genauen Namen des Spiels leider gerade nicht. Die 2. Klasse langweilt sich beim Vorlesen einzelner Buchstaben.

Leseblitz: die SuS lesen teilweise nicht mit, schalten ab. Ich vermute, dass es am (altersabhängigen) Interesse am Thema und dem fehlenden spielerischen Charakter oder dem Lesen selbst (und die Wiederholung des Spieles) liegt.

Übung, bei der die Buchstaben in die richtige Reihenfolge gebracht werden müssen, Übung zum Nachsprechen

Wörtersprint, ABC Profi, Wörtersuche

g) Wo genau gibt es Probleme mit der Durchführung im Team?

manche Kinder streiten sich während des Trainings - manche Kinder konzentrieren sich eher auf den Partner.

Bei der Korrektur

Bei Krankheitsfällen sind Einzelspieler besser zu organisieren

Das gegenseitige coachen funktioniert nicht immer, besonders wenn die Niveaus nicht komplett identisch sind. Teilweise war es besser, wenn ich als Betreuer eingesprungen bin.

Die Gruppen sind teilweise sehr heterogen. Daher macht das Spielen im Single-Modus mehr Sinn. Selten sind Kinder genau gleich eingestuft und es ist nicht ganz fair, wenn einige Kinder in Teams spielen dürfen und andere alleine vor den Tablets sitzen. Die Einteilung ist eine große Herausforderung in der Praxis.

Die Teamspieler können ab und zu Stress miteinander haben oder einander ablenken. Bei manchen Teams war am Anfang die Geschlechterkonstellation ein Problem, besonders im fortgeschrittenem Alter.

Es gibt Probleme, wenn man die Kinder ohne Absprachen in Teams spielen lässt. Es ist wichtig, mit den Kindern über das Arbeiten im Team zu sprechen und gemeinsame Regeln zu vereinbaren, damit es nicht zu Streitigkeiten im Team kommt.

Es ist immer abhängig von den Kindern. Manche Kinder lernen gern gemeinsam im Team und andere lernen effektiver wenn sie allein arbeiten.

Es ist nicht immer einfach, gute Teams zu finden. Man muss die Lesekompetenz (das schwächere Teammitglied verliert evtl. Motivation durch Vergleich), die Verträglichkeit der SuS miteinander (gute Verträglichkeit: SuS nutzen die Anwendung nicht, weil sie sich eher miteinander beschäftigen, schlechte Verträglichkeit: SuS streiten oder kooperieren nicht) und die zeitliche Verfügbarkeit achten, wenn das Tutoring, wie bei uns, im laufenden Unterricht eingebunden ist (mit DaZ Komponente inkl. der organisatorischen Probleme durch Teilintegration).

Es zeigen sich evtl. später Unterschiede im Lernstand

Falsch klicken- obwohl richtig gelesen wurde, Quatsch machen, einfach weiter klicken ohne zu bearbeiten

Kinder sind oft ungeduldig und teilweise ist es sehr laut

Kinder vergessen zu wechseln oder korrigieren sich nicht konsequent

Machen sich gegenseitig "lustig" über die kindische Art des Programms; verstehen nicht, wer wann an der Reihe ist

siehe 15

siehe Frage vorher, Schummeln beim gegenseitigen Vorlesen und markieren der fehlerhaften Wörter

Sprachprobleme

Unterschiedliche Leistungsstufen und Frustrationsgrenzen. Gegenseitige "Sabotage", Übergenauigkeit bzw. "das Suchen von Fehlern" demotivieren die Spieler und erhöht die Frustration.

V.a. unterschiedliche Motivation der Teammitglieder- die einige Übungen (Buchstaben) nicht mehr ernst nehmen und dadurch die Motivation des Teams sinkt. Immer wieder derselbe Spruch am Ende einer Übung.

Wenn ein Teammitglied wenig motiviert ist, leidet die Aufmerksamkeit. Außerdem kommt es dann eher dazu, dass beide einfach Quatsch machen. Außerdem sind die Kinder nicht immer in der Lage, die/alle Fehler zu finden oder Fragen zu beantworten.

h) Wie schätzen Sie die Teilnahmebereitschaft der Kinder für „Lesen mit dem Turbo-Team“ im Allgemeinen ein? Sonstiges:

Anfangs hoch, später lässt sie ein wenig nach, aber steigt immer wieder hoch

Das ist pauschal nicht zu sagen, bei manchen nimmt die Freude daran mit der Zeit zu, bei machen ab.

habe noch zu wenig Erfahrung

sehr unterschiedlich und individuell