



Einzelfallorientierung in der Evidenzbasierung der Sprachtherapie*

Case orientation in evidence-based practice of speech language therapy

Jürgen Kohler, Anja Starke, Anke Kohmäscher

Zusammenfassung

Ein evidenzbasiertes Vorgehen in der Sprachtherapie wird seit vielen Jahren gefordert. Allerdings fehlt es in vielen Bereichen an den erforderlichen Studien und zusätzlich wird die Übertragbarkeit der Hierarchie von Studiendesigns aus der evidenzbasierten Medizin kritisch diskutiert. In dem vorliegenden Beitrag wird die in der Sprachtherapie wesentliche Orientierung am Einzelfall als Ausgangspunkt genommen, um die Eignung verschiedener Studiendesigns für die Evidenzbasierung in der Sprachtherapie zu diskutieren. Hierzu wird zunächst die Einzelfallstudie in ihren Grundzügen vorgestellt und über verschiedene methodische Varianten ihr Potenzial zu einer größeren Verallgemeinerbarkeit der Ergebnisse dargestellt. Der Übergang von Einzelfallstudien zu versorgungsorientierten Gruppenstudien erscheint dabei fließend. Diese methodischen Überlegungen werden anhand von drei Studien aus dem Bereich der Redefluss- und Kommunikationsstörungen konkretisiert. Abschließend werden die methodischen Erkenntnisse aus diesen Studien zusammengeführt und die Implikationen für die Evidenzbasierung in der Sprachtherapie herausgestellt.

Schlüsselwörter

Evidenzbasierung, Einzelfallorientierung, Einzelfallstudien, Versorgungsforschung, Sprachtherapie, Logopädie

Abstract

Since several years there is a claim for evidence-based practice in speech language therapy. However, in many areas there is still a lack of studies. Additionally, some authors argue that the hierarchy of study designs in evidence-based medicine should not simply be transferred to speech language therapy. In this article case orientation that is crucial in speech language treatment is taken as a starting point to discuss the suitability of different study designs for evidence-based speech language therapy. At first, we explain the fundamentals of single subject (case) studies and some methodological variations that have the potential for better generalization. Doing this, we see a fluent transition from case studies to comparative effectiveness research. The methodological considerations are then illustrated by three studies in the field of fluency and communication disorders. Finally, the methodological findings from these studies are taken together to discuss implications for evidence-based speech language therapy.

Keywords

evidence-based practice, case orientation, single subject studies, health services research, speech language therapy

* Dieser Beitrag hat das Peer-Review-Verfahren durchlaufen.

1 Einleitung

Seit geraumer Zeit wird über die Notwendigkeit der Evidenzbasierung von logopädischen bzw. sprachtherapeutischen¹ Interventionen diskutiert (z. B. Beushausen, 2014, Borgetto et al., 2016). Während randomisierte, kontrollierte Studien in der Medizin als der Goldstandard gelten, wird Einzelfallstudien in der sogenannten Evidenzhierarchie eine eher untergeordnete Rolle zugestanden (Yin, 2014). Die Übertragbarkeit dieser Stufen aus der evidenzbasierten Medizin (Abb. 1) auf die Gesundheitsberufe wird durchaus kritisch gesehen und teilweise in Frage gestellt (Kraus, 2018). Ein wesentlicher Grund hierfür ist die schlechte Übertragbarkeit von Studienergebnissen auf die eigene klinische Praxis (Borgetto et al., 2016). Experimentelle Studien, die in der EBP-Hierarchie weit oben angesiedelt werden, weisen eine hohe interne Validität auf, indem Störeinflüsse u.a. über Randomisierung, eine Kontrollgruppe und strikte Einschlusskriterien für die teilnehmenden Personen minimiert werden. Das alleinige Zurückführen der Wirkung auf die Intervention wird unterstützt durch standardisierte und manualisierte Vorgehensweisen. Damit entspricht das Ergebnis dieser sogenannten klinischen Studien jedoch wenig dem Versorgungsalltag, der in der Sprachtherapie durch komplexe, individualisierte Interventionen sowie heterogene Patientengruppen geprägt ist. Zu Recht beschreiben Borgetto et al. (2016) die Forderung nach „Evidenzen, [die] für die therapeutische Praxis brauchbar sind.“ (S. 25).

| Evidenzstufen | | |
|-----------------------|---|--|
| ↑ Güte der Evidenz | Reviews/Metaanalysen zusammenfassende Übersichten | Verringerung des Einflusses weitere Zufallsfaktoren (z. B. Rekrutierung, Therapeut, Stichprobe etc.) |
| | Experimentelle Studien Kontrollgruppe, Randomisierung | Ausschaltung von systematischen Unterschieden zwischen Gruppen |
| | Quasi-experimentelle Studien Kontrollgruppe | Kontrolle über nicht therapiebedingte Veränderungen (z. B. Entwicklung) |
| | Verlaufsstudie ohne Kontrollgruppe Gruppenmittelwerte | Zufallsfaktoren haben weniger Einfluss auf das Ergebnis |
| | Einzelfallstudien viele Zufallsfaktoren | Objektive Messung der Veränderung (z. B. mittels Tests) |
| | Expertenmeinungen (Praktiker) Subjektivität | Praxiserfahrung |

Abb. 1: Klassische Evidenzhierarchie mit steigender Güte der Evidenz von unten nach oben

Therapeutisch verwertbare Evidenzen scheinen eher aus Einzelfall- und Beobachtungsstudien gewonnen werden zu können. Aus diesem Grund schlagen Borgetto et al. (2016) eine adaptierte EBP-Hierarchie, nämlich die Forschungspyramide, vor. In dieser wird die klinisch-experimentelle Forschung um die Ebenen der qualitativen Forschung als auch der Versorgungsforschung ergänzt (Kraus, 2018).

Ziel des vorliegenden Artikels ist es, die Bedeutung der Einzelfallorientierung in der Evidenzbasierung der Sprachtherapie herauszuarbeiten. Da die klassische Einzelfallstudie auch in der Forschungspyramide an unterster Stelle angesiedelt ist, werden neben einer Darstellung der Grundzüge auch methodische Varianten der Einzelfallstudie dargestellt, die eine höhere Beweiskraft und Verallgemeinerbarkeit ermöglichen.

Diese theoretischen Ausführungen werden im Anschluss veranschaulicht durch die exemplarische Darstellung von drei einzelfallorientierten Studien im Kontext der Redefluss- und Kommunikationsstörungen. Diese zeigen mit verschiedenen Schwerpunkten auf, wie eine Einzelfallorientierung bei unterschiedlichen Designs und Fragestellungen gelingen kann.

Anhand dieser Erfahrungen im Bereich der Redefluss- und Kommunikationsstörungen soll eine Diskussion angeregt werden, inwiefern Einzelfallstudien auch in anderen Themenbereichen zur Evidenzbasierung in der Sprachtherapie beitragen können.

1 Im Folgenden nutzen wir die Begriffe Sprachtherapie und sprachtherapeutisch. Dabei schließen wir jeweils alle Berufsgruppen mit ein, die sprachtherapeutisch bzw. logopädisch tätig sind.

2 Einzelfallstudien in den Sozialwissenschaften

(Einzel-)Fallstudien können in unterschiedlicher Form angelegt sein (Petermann, 1996a, 1996b). Ganz grob wird in eine *qualitative Einzelfallstudie* (Flick et al., 2012) und eine *kontrolliert-quantitative Einzelfallstudie* (Julius et al., 2000; Reicherts & Genoud, 2015) differenziert. Letztere wird meist auch nur kontrollierte Einzelfallstudie genannt, ist aber klar dem sogenannten quantitativen Paradigma der Sozialwissenschaft verpflichtet (Döring & Bortz, 2016).

2.1 Die qualitative Einzelfallstudie

Das qualitative Paradigma der Sozialwissenschaft ist an sich eng mit der Methode der Fallstudie verknüpft. Der Fall betrifft dabei nicht notwendigerweise nur eine Person, sondern kann als eine Einheit von mehreren Personen begriffen werden, die sich in ganz bestimmten Merkmalen gleichen. So kann eine sprachtherapeutische Praxis mit den dort arbeitenden Fachkräften oder eine einzelne Sprachtherapie mit den beteiligten Personen (Klientel, Bezugspersonen, therapeutische Fachperson) als ein Fall betrachtet werden.

An dieser Beschreibung des Falles sieht man, dass unterschiedliche Abstraktionsmöglichkeiten der Fallbetrachtung (z. B. Mikro-, Meso-, Exo- und Makrosystem) möglich sind (Kasten, 2014). Einzelfälle können auf allen Systemebenen definiert werden, wobei das Wesen der qualitativen Einzelfallstudie immer die systemüberschreitende Betrachtung ist. Es wird z. B. versucht, das Individuum (Mikroebene) in seinem familiären Kontext (Mesoebene) und auch in seiner Eingebundenheit in eine Institution wie der Schule (Exosystem) zu verstehen. Qualitative Einzelfallstudien sind ihrem Wesen nach also ganzheitlich.

Dieses Wesensmerkmal der qualitativen Einzelfallstudie spiegelt sich auch in den Methoden wider, welche in der Sozialforschung genutzt wird. Bei der qualitativen Datenerhebung sind prinzipiell viele offene Methoden und Techniken wie teilnehmende Beobachtung, Interviews, Gruppendiskussionen oder Inhaltsanalysen von Dokumenten denkbar (Lamnek & Krell, 2016).

Datenauswertung und -interpretation sind in der qualitativen Forschung schwer zu trennen und haben die Aufgabe, ein Kondensat des ganzheitlichen Verstehens in Form von nachvollziehbaren Kategorien zu erzeugen. Diese haben zunächst nicht den Anspruch verallgemeinerbar zu sein und gelten nur für den untersuchten Einzelfall. Auf die Situation der Sprachtherapieforschung bezogen bedeutet das: Die Kategorien sind als Hauptaussagen des subjektiven Empfindens und Erlebens von Klienten, Bezugspersonen und therapeutischen Fachkräften zu betrachten. Qualitative Einzelfallforschung bemüht sich aber um intersubjektive Nachvollziehbarkeit, indem z. B. die Daten eines Interviews von zwei unabhängig voneinander agierenden Personen ausgewertet und interpretiert werden und bei fehlender Übereinstimmung nach den Gründen dafür gesucht wird.

2.2 Die quantitativ-kontrollierte Einzelfallstudie

Im Gegensatz zur qualitativen Einzelfallstudie konzentriert sich die quantitativ-kontrollierte Einzelfallstudie auf einzelne Personen und einzelne Variablen. In Analogie zum experimentellen Setting werden dabei wenige Variablen aus dem komplexen Verhalten und Erleben ihrer Versuchspersonen isoliert. Die Abbildung 2, welche von Jain und Spieß (2012) entlehnt ist, soll diese Konzentration veranschaulichen.

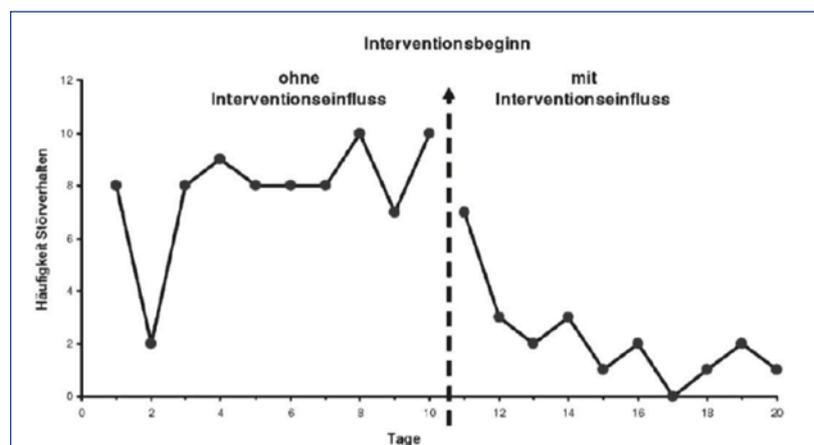


Abb. 2: Prototypischer Aufbau einer einfachen quantitativ-kontrollierten Einzelfallstudie (aus Jain & Spieß, 2012, S. 214)

Kennzeichnend für quantitativ-kontrollierte Einzelfallstudien ist die Unterscheidung von A- und B-Phasen (Kulawiak & Wilbert, 2018). Man spricht von der Phase A (Baseline-Phase), wenn während dieses Zeitraumes keine Interventionen stattfinden. Die Phase mit Interventionen wird B-Phase benannt. Anstelle vieler Probanden mit zwei Messzeitpunkten (prä-post) werden in Einzelfallstudien in jeder Phase mehrere Messungen der abhängigen Variable (in Abbildung 2 die Häufigkeit von Störverhalten) durchgeführt, was auch als Messreihe bezeichnet wird. Die maßgebliche Variation der abhängigen Variable entsteht durch die Intervention, die in der B-Phase eingeführt wird. Wenn bei Einführung einer Intervention eine deutliche Veränderung der abhängigen Variable sichtbar wird, können daraus Rückschlüsse auf Ursache-Wirkungszusammenhänge gezogen werden. Erhärtet werden können diese durch ein Absetzen der Intervention in einer anschließenden A-Phase oder durch die Einführung einer weiteren Intervention (C-Phase).

Das Grundprinzip der quantitativen Einzelfallstudie mit einer Abfolge von A- und B-Phasen ist beliebig variierbar (vgl. 2.3). Dabei gehen die unterschiedlichen Variationen der Phasenkombination auf unterschiedliche Fragestellungen zurück (vgl. dazu Reicherts & Genoud, 2015).

Die Datenerhebung der quantitativen Einzelfallstudie ist demensprechend geprägt von der Messbarkeit des Verhaltens. Es werden Häufigkeiten gezählt oder Leistungen skalierend eingeschätzt (z. B. niedrig-mittel-hoch) bzw. mit Zahlen belegt wie 0 für keine Leistung und 100 für maximale Leistung. Zum Einsatz kommen dabei standardisierte Verfahren wie zum Beispiel Fragebögen mit geschlossenen Antwortformaten und Beobachtungen mit genau vorab festgelegten Skalen. Die Auswertungsmethoden sind ebenfalls von der Quantifizierung geprägt. Über visuelle Inspektion wird der Verlauf einer Leistung deskriptiv dargestellt (wie oben durch ein Liniendiagramm) oder auch inferenzstatistisch auf Signifikanz überprüft. Mit letzterem wird versucht, einen gefundenen Ursachen-Wirkungs-Zusammenhang von einem zufällig entstandenen Zusammenhang abzugrenzen.

Wie oben schon erwähnt, wird diese Form der Einzelfallstudie auch kontrollierte Einzelfallstudie genannt. Das trifft das Wesen des Vorgehens sehr gut, da es im Kern um die Kontrolle der Variablen geht, um so reliabel wie möglich und frei von sogenannten Störvariablen einen Ursachen-Wirkungs-Zusammenhang mit hoher Wahrscheinlichkeit nachweisen zu können. In der Interpretation der Daten wird dabei auf eine zuvor aufgestellte Hypothese zurückgegriffen, die sich z. B. auf die Wirksamkeit einer Intervention beziehen kann. Subjektiv ist diese Form von Evidenzforschung, weil sich der vermeintlich belegte Ursachen-Wirkungs-Zusammenhang nur auf ein Subjekt bzw. einen Fall bezieht. Die Verallgemeinerbarkeit über die in einem Fall agierenden Subjektkonstellation (Klientel Bezugspersonen, therapeutische Fachpersonen) hinaus ist schwierig und damit die externe Validität im Sinne einer Verallgemeinerbarkeit auf andere Personen (Döring & Bortz, 2016) gering.

2.3 Varianten der quantitativ-kontrollierten Einzelfallstudie

Wie bereits erwähnt, lässt die Kombination von A- und B-Phasen vielfältige Varianten der quantitativ-kontrollierten Einzelfallstudie zu. Beispielsweise verzichtet die *kontrollierte Einzelfallstudie im psychometrischen Ansatz* auf eine Messreihe über die Phasen hinweg und setzt normierte Testverfahren vor und nach der Intervention ein. Statt des Vergleichs zwischen den Messzeitpunkten stellt die Normierungsstichprobe der Tests eine Vergleichsebene für den Einzelfall dar. Dieser Normbezug bringt die für die statistische Auswertung so wichtige Varianz und damit die Möglichkeit, die Intervention als nicht nur zufällige, sondern systematische Wirkung auf die abhängige Variable zu identifizieren. Der psychometrische Ansatz hat einen wichtigen Vorteil. Er ist ökonomisch und wahrscheinlich praxisnah. Die sprachtherapeutische Fachperson, welche ihre Eingangs- und Abschlussdiagnostik mit Hilfe von normierten Tests durchführt, erfüllt die Anforderungen dieses Designs. Außerdem liefert ein Test i.d.R. eine Variablenvielfalt innerhalb des interessierenden Konstrukts. So bestehen Tests, die das Konstrukt Sprachentwicklung messen, in der Regel aus vielen Items, die innerhalb des Konstrukts dessen Unterkategorien (z. B. Grammatik, Semantik oder Phonologie) in ihrer Ausprägung bestimmen.

Das *Multiple-Baseline-Design* ist ebenfalls eine vielgenutzte Variante und unterscheidet sich insbesondere durch die Variablenvielfalt, da mehrere abhängige Variablen (AV) und damit einhergehende unabhängige Variablen (UV) Gegenstand der Einzelfallstudie sind. Im Prinzip wird dabei ein AB-Plan auf unterschiedliche Verhaltensweisen oder Problembereiche angewendet (Jain & Spieß, 2012). Die meisten Einzelfallstudien im Multiple-Baseline-Design gehen dabei mit einer Nacheinanderlogik der Experimental-Variablenreihung vor. Das heißt, dass die Interventionen (= UVs) nicht gleichzeitig durchgeführt werden, sondern nacheinander. Dieses forschungs-

methodische Vorgehen soll die Vermischung der Effekte von unterschiedlichen Interventionen und ihre gegenseitige Störung auf die Identifikation des Ursachen-Wirkungs-Zusammenhangs verhindern (Kulawiak & Wilbert, 2018).

Eine weitere Variante des Multiple-Baseline-Designs, welches wir zur besseren Abgrenzung im Folgenden als *kumulierte Einzelfallstudie* bezeichnen, ist charakterisiert durch die Variablenvielfalt auf Personenebene. Dabei wird die gleiche Intervention bei unterschiedlichen Personen untersucht. Dies kann parallel oder auch nachträglich geschehen. Dies setzt eine gewisse Vergleichbarkeit hinsichtlich der Fragestellung und Studiendurchführung voraus. Gewollte Variabilität herrscht dagegen in allen Merkmalen der Fälle, wie z. B. der sozio-ökonomische Status oder das Geschlecht der Klientel, die Berufserfahrung der therapeutischen Fachpersonen oder das Alter der Bezugspersonen. Durch die Anhäufung der Einzelfälle besteht die Chance, dass sich solche spezifischen Merkmale über die Betrachtung von vielen Fällen hinweg nivellieren und der reine Effekt der therapeutischen Maßnahme sichtbar wird. In der Kumulation steckt also die Chance der Verallgemeinerbarkeit (Kulawiak & Wilbert, 2018). Es kann vom Einzelfall abstrahiert werden und die Wirkung einer Intervention für viele Fälle dargestellt und gegen zufällige oder störende Effekte abgegrenzt werden. Dieses Potential der kumulierten Einzelfallstudie entspricht dem Anspruch der Evidenzforschung, wenn sie Therapieansätze, -modelle oder -programme auf ihre Wirksamkeit hin prüfen will.

Weitere Variationen im Studiendesign einer kontrollierten Einzelfallstudie sind über die Kombination der oben genannten Designs denkbar und in Anwendung. Variablenvielfalt kann beispielsweise auch innerhalb der Interventionen bestehen. So können praxisnahe Variationen von Therapieansätzen trotz einer gewissen Flexibilisierung in der Ausgestaltung durch die therapeutische Fachperson oder trotz einer Anpassung an die individuellen Bedürfnisse des Klienten verglichen werden. Entscheidend ist die Wirkung (= Veränderung der abhängigen Variable/n) und diese wird durch den Vergleich etwa zwischen A- und B-Phase oder den zwei Testzeitpunkten relativ unabhängig vom Zufall gemessen. Ebenso sind Kombinationen von quantitativen und qualitativen Methoden im Sinne eines mixed-method-Designs denkbar, um die Vorteile beider Methodenbereiche zu kombinieren.

2.4 Kumulierte Versorgungsstudien

Eine kumulierte Einzelfallstudie ist strukturell sehr einfach in eine Gruppenstudie überführbar, indem man zwei Gruppen von Einzelfällen bildet und bei diesen Gruppen unterschiedliche Interventionen durchführt. Dies verdeutlicht den fließenden Übergang von Einzelfallstudien zu einzelfallorientierten Versorgungsstudien, die in der Forschungspyramide als quantitativ oder qualitativ ausgerichtete Beobachtungsstudien bezeichnet werden (Borgetto et al., 2016). Die Einzelfallorientierung deutet dabei an, dass auch eine Gruppenstudie letztendlich aus Einzelfällen besteht. Jede einzelne Intervention findet dabei vor dem Hintergrund komplexer Kontextbedingungen statt, welche die Wirkung der Intervention mit beeinflussen (Pfaff et al., 2017). Aus Sicht der Versorgungsforschung entspricht dies dem Versorgungsalltag, wodurch Ergebnisse unmittelbar auf das tatsächliche Versorgungsgeschehen übertragbar sind.

Die Komplexität von Interventionen und Kontexten, in der Versorgungsforschung als doppelte Komplexität bezeichnet, ist auch für Forschende eine Herausforderung und erfordert ein zusätzliches Augenmerk auf qualitative Forschungsmethoden (Meyer & Flick, 2017). Ausgehend von einer Patienten- und Populationsperspektive treten dabei patientenrelevante Endpunkte wie Lebensqualität, Aktivität und Teilhabe immer mehr in den Vordergrund (Gräßel et al., 2015). Diese lassen sich insbesondere über eine Kombination von Strategien der quantitativen Sozialforschung mit denen der qualitativen Sozialforschung erheben (mixed-methods designs). Der Mehrwert von qualitativen Strategien, welche die Einzelfallanalysen ergänzen, liegt in der Entdeckung von bisher unbedachten oder unbekanntem Variablen. Hierzu ein Beispiel: Interventionsvorschläge im sprachtherapeutischen Feld haben neben der sprachlichen Ebene im engeren Sinne auch immer allgemein didaktische oder formal-gestalterische Elemente zum Gegenstand, welche ihren Einsatz prägen. So kann die Attraktivität von Zeichnungen oder Bildern bei Materialien für die Sprachtherapie Einfluss auf die Wirksamkeit einer Intervention haben, welche mit dem Material verbunden ist. Wenn nun viele Anwender hinsichtlich der Qualität von Zeichnung oder Bildern eine Rückmeldung geben, kann dies ein Hinweis auf deren Optimierungsbedarf sein (bei kritischen Rückmeldungen) oder auch eine Bestätigung ihres positiven Wertes für den Effekt des Materialeinsatzes. Qualitative Rückmeldungen haben insofern ein hypothesengenerierendes Potential. Eine bis dahin nicht explizierte Frage wird aufgeworfen und kann genauer unter die Lupe genommen werden.

Aufgrund des Anspruchs der Verallgemeinerbarkeit in der Evidenzforschung kommt der Kumulation von Einzelfällen für die Evidenzbestimmung von Therapieansätzen, -modellen oder -programmen ein hohes Gewicht zu. Welche der anderen Varianten einer kontrollierten Einzelfallstudie zum Zuge kommen, ist eher eine Frage der Ökonomie bzw. Durchführbarkeit an sich. Oben wurde schon angedeutet, dass viel für den psychometrischen Ansatz spricht. Allerdings wird es im Zuge des Einsatzes von digitaler Technik immer realistischer, dass auch Messreihen mit einer gewissen Variablenvielfalt wie beim multiple-baseline-design zum Zuge kommen können. So kann z. B. das Smartphone eingesetzt werden, um Klienten oder Bezugspersonen in ihrem Alltag aufzufordern (auch mehrmals am Tag), subjektive Bewertungen über den sprachlichen Leistungsstand abzugeben. Im Folgenden zeigen wir drei Varianten einzelfallorientierter Studie im Kontext von Redefluss- und Kommunikationsstörungen.

3 Einzelfallstudien mit multiple-baseline-Design zur Evaluation der Dortmunder Mutismustherapie

Die Dortmunder Mutismus-Therapie (DortMuT) ist ein in sich interdisziplinäres Therapiekonzept, in das durch seine Entwicklungsgeschichte bereits von Beginn an psycho- und sprachtherapeutische Methoden eingeflossen sind (Subellok et al., 2012). Das vorrangige Ziel der Therapie ist es, schweigende Kinder und Jugendlichen selbstbestimmt ins Sprechen zu begleiten. Dazu werden ihnen Interaktionsangebote gemacht, die sie annehmen können, jedoch nicht müssen. Bei Kindern im Vorschulalter geht es vor allem um das Erlernen neuer interaktiver Kompetenzen im Spiel. Im Schulalter wird zunehmend direkter gearbeitet und der Aufbau des sprachlich-kommunikativen Verhaltens mit sprach- und verhaltenstherapeutischen Methoden unterstützt.

Langfristiges Ziel einer jeden Therapie ist es, dass die Kinder und Jugendlichen selbstbestimmt sprechen und ihren Alltag kommunikativ meistern können. Vor diesem Hintergrund ist nicht allein die Quantität des Sprechens ein Maß für den Erfolg einer Therapie im Sinne von DortMuT. Vielmehr geht es um die Qualität des Sprechens. Denn auch der Inhalt der Kommunikation, situative Aspekte und auch die anwesenden Personen können die Interaktion von Kindern und Jugendlichen mit selektivem Mutismus deutlich beeinflussen (Melfsen & Walitza, 2017). So kann das Vorlesen eines Textes aus einem Schulbuch vielleicht schon gut klappen, das Vorlesen der eigenen Hausaufgaben jedoch noch nicht, da es sich hier um persönliche Informationen handelt. Auch könnte es sein, dass eine Person bereits ausführlich auf Fragen anderer Personen sprachlich reagiert, es ihr aber noch nicht gelingt, eigenaktiv eine Interaktion zu beginnen oder eigene Ideen einzubringen. Somit ist eine zentrale Frage bei der Evaluation von DortMuT vor allem die Frage nach dem Maß für den Erfolg einer Therapie.

3.1 Therapieforschung zu selektivem Mutismus

Die Therapieforschung im Bereich des selektiven Mutismus steht aktuell noch eher am Anfang. Insbesondere für die gerade im deutschsprachigen Raum prominenten sprachtherapeutischen Konzepte stehen Evaluationsstudien noch weitestgehend aus. Im internationalen Raum lassen sich bereits zahlreiche Therapiestudien finden (für einen Überblick siehe etwa Anstendig, 1998; Cohan et al., 2006; Pionek Stone et al., 2002; Standart & Le Couteur, 2003; Viana et al., 2009; Zakszeski, 2016). Diese befassen sich jedoch maßgeblich mit verhaltenstherapeutischen Konzepten und weisen auf die Wirksamkeit von kognitiv-behavioralen Behandlungsmethoden bei selektivem Mutismus hin. Bezogen auf das Studiendesign wurde bislang bereits in zahlreichen Fällen auf mehr oder weniger kontrollierte Einzelfallstudien im Bereich des selektiven Mutismus zurückgegriffen. Dies ist vermutlich auf die geringe Prävalenz und große Heterogenität der Klientel zurückzuführen. Einige wenige Forschungsgruppen haben kontrollierte Gruppenstudien umgesetzt (Bergman et al., 2013; Oerbeck et al., 2014; Oerbeck et al., 2015).

Innerhalb der Vielzahl von Einzelfall- und kleineren Gruppenstudien wurde bislang maßgeblich mit Prä-Posttest-Designs gearbeitet. Als AV wurde in der Regel das Sprechverhalten der Kinder bzw. Jugendlichen in unterschiedlichen Kontexten erfasst, wobei meist Fremdeinschätzungen durch Eltern und Lehrkräfte genutzt wurden. Differenzierte Maße zur Qualität der Interaktionen sind nur selten im Fokus. Zusätzlich greifen viele Forschungsgruppen aufgrund des starken Zusammenhangs von selektivem Mutismus und Angst auf Angstfragebögen als zusätzliches Outcomemaß zurück (siehe dazu Zakszeski, 2016).

3.2 Einzelfallstudien DortMuT

Es wurden bereits mehrere Einzelfallstudien im therapeutischen und schulischen Kontext zur Evaluation von DortMuT durchgeführt, wobei vorrangiges Ziel dieser Studien die Überwindung zahlreicher methodischer Schwierigkeiten waren. Ein wichtiges Anliegen war es dabei, nicht nur die Quantität des Sprechens – also Häufigkeit des Sprechens oder die Äußerungslänge – der Kinder als Outcomemaß zu nutzen, sondern vor allem auch qualitative Aspekte – z. B. unterschiedliche Personen, Situationen, Inhalte, Möglichkeiten der Bedürfnisäußerungen – als Maße für einen Therapieerfolg mit einzubeziehen. Im Folgenden werden zwei Einzelfallstudien, eine aus dem schulischen, eine aus dem therapeutischen Kontext vorgestellt. Ein Schwerpunkt liegt dabei auf den methodischen Aspekten und deren Umsetzbarkeit im Feld.

3.2.1 Schulische Kurzzeitintervention

Im Rahmen einer Abschlussarbeit wurde eine schulische Kurzzeitintervention bei einem schweigenden Zwillingsspaar im Alter von 8;11 Jahren begleitet und analysiert (Starke et al., 2018). Dabei wurde ein A-B-C-Design verfolgt (siehe auch Abb. 3), wobei innerhalb der A-Phase (Baseline) über drei Wochen das Sprech- und Interaktionsverhalten der Kinder im Unterricht und in familiären Kontexten mit maßgeblich quantitativen Methoden erfasst wurde. Anschließend erhielten die Kinder eine sechswöchige Intervention innerhalb der Schule. In der Follow-Up-Phase (C-Phase) wurden an zwei Messzeitpunkten (kurz nach der Intervention und nach knapp 3 Monaten) Daten zum Sprech- und Interaktionsverhalten der Kinder in schulischen und außerschulischen Kontexten erhoben.



Abb. 3: Studiendesign der Einzelfallstudie im schulischen Kontext

Im Elternfragebogen wurden anhand von knapp 60 Items mutismusspezifische Verhaltensweisen in unterschiedlichen, von Eltern beurteilbaren Situationen erfasst. Die Eltern schätzten auf einer fünfstufigen Skala an, wie sehr das beschriebene Verhalten auf das Kind zutrifft. Hohe Werte im Elternfragebogen spiegeln das Vorliegen mutismustypischer Verhaltensweisen wider. Im schulischen Kontext wurden unterschiedliche Informationsquellen genutzt, um ein möglichst valides Bild über die Entwicklung der Kinder zu erhalten. Mithilfe des Dortmunder Mutismus Screenings (DortMuS-Schule, Starke & Subellok, 2016) wurden in der Baseline- und Follow-up Phase Informationen zum Sprechverhalten und Möglichkeiten der Bedürfnisäußerung der Kinder über die Lehrkräfte erhoben. Anhand von strukturierten Unterrichtsbeobachtungen wurden ebenfalls vor und nach der Intervention Daten zum Interaktions- und Meldeverhalten der Kinder im Unterricht erfasst. Die Tagebuchdokumentation diente darüber hinaus dazu, engmaschig das Sprech- und Meldeverhalten der Kinder im Unterricht zu dokumentieren. Dabei wurden die Lehrkräfte der Kinder aufgefordert, an mindestens 3 Tagen in der Woche in der jeweils dritten und vierten Unterrichtsstunde Strichliste zu verschiedenen Aktivitäten der Kinder (z. B. *Kind meldet sich im Unterricht* oder *Kind sagt ein Wort*) zu führen. Die Tagebuchdokumentation erfolgte in der Baseline-, Interventions- und Follow-up-Phase.

Ein kurzer Überblick zu den Ergebnissen der Studie findet sich in Tabelle 1. Sowohl innerhalb der Fragebögen der Lehrkräfte als auch der Eltern zeigen sich deutliche Hinweise auf Veränderungen des Sprech- und Interaktionsverhaltens der Kinder. Kind 1 schwieg bereits vor der Intervention deutlich weniger als Kind 2. Entsprechend ergaben sich auch auf der Skala Sprechen und Schweigen, welche das Sprechverhalten des Kindes in unterschiedlichen schulischen Situationen fokussiert, nur geringe Veränderungen. Unterschiede sind vor allem im Bereich Bedürfnisäußerung, also in eher inhaltlich-situativen Aspekten des Interaktionsverhaltens auszumachen. Bei

Kind 2 hingegen ist eher eine Abnahme des schweigenden Verhaltens sichtbar, wenngleich die Ergebnisse von DortMuS-Schule weiterhin auf eine deutliche Symptomatik des selektiven Mutismus hinweisen. Dies spiegelt sich auch in den Unterrichtsbeobachtungen wider. Auch hier zeigte sich eine deutliche Zunahme verbalen Interaktionsverhaltens bei Kind 1. Kind 2 zeigte vor allem Veränderungen im nonverbalen Verhalten bei weiterhin deutlichen Anzeichen eines selektiven Mutismus.

Die Tagebuchdokumentationen, die bei der Studienkonzeption als wichtigste Datenbasis zur Auswertung gedacht waren, war bei beiden Kindern nur wenig aussagekräftig. Bei Kind 2 fiel kurz nach Beginn der Studie die Klassenlehrkraft krankheitsbedingt aus und es gab keine kontinuierliche Vertretung, die die Dokumentation des Verhaltens übernahm. Kind 1 zeigte Fortschritte im Interaktionsverhalten vor allem im Fachunterricht und nicht im Unterricht der Klassenlehrkraft, welche die Tagebuchdokumentation übernommen hatte.

Tab. 1: Übersicht über die wichtigsten Ergebnisse der Einzelfallstudie

| | Kind 1 | | Kind 2 | |
|----------------------------------|------------------|------------------|------------------|--------------------|
| | Baseline | Follow-up | Baseline | Follow-up |
| DortMuS-Schule (gesamt) | 41 | 27 | 63 | 49 |
| Sprechen/Schweigen | 21 | 17 | 37 | 26 |
| Bedürfnisaußerung | 20 | 10 | 26 | 23 |
| Tagebuchdokumentation | | | | |
| Sprech- u. Meldeverhalten | Stark schwankend | | Nicht auswertbar | |
| Unterrichtsbeobachtung | | | | |
| Interaktionsverhalten nonverbal | | Unverändert viel | | Mehr nonverbale K. |
| Interaktionsverhalten verbal | | Mehr Sprechen | | Wenig Sprache |
| Elternfragebogen (gesamt) | 47 | 29 | 72 | 36 |

Anmerkungen. Die Werte von DortMuS-Schule sowie dem Elternfragebogen sind aufsummierte Werte. Hohe Werte spiegeln typisches Verhalten für einen selektiven Mutismus in vielen Situationen wider.

Die Erfahrungen aus der schulischen Interventionsstudie zeigen die große Komplexität der Evaluation einer Mutismustherapie auf. Da die Kernsymptomatik des selektiven Mutismus stark durch personelle und situative sowie inhaltliche Aspekte beeinflusst sein kann und man häufig zu Beginn einer Therapie nicht abschätzen kann, in welchen Kontexten das Kind zuerst kommunikative Fortschritte vollziehen wird, bedarf es insbesondere bei einer derartigen Kurzzeitintervention einer Dokumentation zahlreicher Kontexte. Allerdings stehen meist nicht die (personellen usw.) Ressourcen zur Verfügung, um alle potenziell bedeutsamen Kontexte eines Kindes zu dokumentieren.

3.2.2 Ambulante Sprachtherapie

Im Rahmen einer Einzelfallstudie innerhalb einer ambulanten sprachtherapeutischen Behandlung war es das Ziel, den Einfluss dieser auf die für den Alltag bzw. therapeutischen Fortschritt des Kindes relevanten Kontexte Familie, Kindertageseinrichtung (Kita) und Sprachtherapie zu untersuchen. Auch in dieser Studie wurde ein multiple-baseline-Design in Kombination mit psychometrischen Testverfahren umgesetzt. Dabei wurden Informationen zum Sprech- und Interaktionsverhalten in einer fünfwöchigen Baselinephase, einer Interventionsphase von insgesamt 15 Wochen sowie einer Follow-Up-Phase erfasst (Abbildung 4).

4 Versorgungsstudie PMS KIDS zur Wirksamkeit von Stottertherapie unter Alltagsbedingungen

Wie in Kapitel 1 beschrieben können Einzelfallstudien in sehr unterschiedlichen Varianten umgesetzt werden. Im Folgenden wird die Studie PMS KIDS vorgestellt, die als Versorgungsstudie zur Evidenzbasierung in der Stottertherapie beitragen soll, dabei aber auch eine ausgeprägte Einzelfallorientierung aufweist.

4.1 Evidenzlage zu Stottertherapien

Die S3-Leitlinie zu Redeflussstörungen, die evidenz- und konsensbasiert entwickelt wurde, zeigt, dass zur Wirksamkeit von Stottertherapien durchaus vielfältige und hochwertige Evidenzen vorliegen. Laut aktueller Evidenzlage wird daher insbesondere die Lidcombe Therapie im Vorschulalter und ab dem Jugendalter eine Stottertherapie nach dem Fluency Shaping Ansatz empfohlen (Neumann et al., 2016). Auffällig ist allerdings, dass für stotternde Kinder im Grundschulalter so gut wie keine Evidenzen vorliegen. Dies ist umso bedeutender, da die Wahrscheinlichkeit für eine Remission in dieser Altersgruppe beträchtlich abfällt und nur noch etwa 50 % beträgt (Natke & Kohmäscher, 2020).

Laut einer Umfrage werden stotternde Grundschul Kinder in Deutschland überwiegend in ambulanten, sprachtherapeutischen Einrichtungen versorgt (Kohmäscher, 2017). Die dort stattfindenden Stottertherapien können laut Heilmittelkatalog bis zu 50 Therapieeinheiten im Regelfall umfassen und finden ein- bis zweimal wöchentlich statt. Sprachtherapeuten setzen verschiedene Therapieverfahren ein, wobei das Stottermodifikationsverfahren KIDS (Kinder Dürfen Stottern, Sandrieser & Schneider, 2015) von einer deutlichen Mehrheit an therapeutischen Fachkräften eingesetzt wird (Kohmäscher, 2019).

Die skizzierte Versorgungssituation zeigt, dass eine Wirksamkeitsüberprüfung von *KIDS* hochrelevant, jedoch auch äußerst schwierig ist. Therapeutische Fachpersonen (n = 90) betreuten zum Zeitpunkt der Befragung durchschnittlich etwa 4-5 stotternde Personen unterschiedlichen Alters und nehmen nur alle 3-4 Monate neue stotternde Personen auf. Zudem wird das Therapieverfahren *KIDS* sowohl von Autoren als auch therapeutischen Fachpersonen als stark einzelfallorientiert und in der Anwendung komplex beschrieben. Für das Projekt PMS KIDS erwies sich der Ansatz der Versorgungsforschung als idealer Weg, um der einzelfallorientierten Herangehensweise von *KIDS* im therapeutischen Alltag Rechnung zu tragen und gleichzeitig methodisch hochwertige Evidenzen zu generieren.

4.2 Studiendesign von PMS KIDS

PMS KIDS ist das Akronym für eine prospektive, multizentrische Studie zur Überprüfung der Wirksamkeit der best-practice Therapie *KIDS*. Diese seit 2018 laufende Studie wird vom Gemeinsamen Bundesausschuss gefördert und erfolgt in Kooperation der FH Münster (Prof. Kohmäscher) mit der Uniklinik Aachen (Prof. Heim). Ziel ist es, die S3-Leitlinie zu Redeflussstörungen in Bezug auf die Altersgruppe der Grundschul Kinder zu ergänzen (Neumann et al., 2016). Als Studiendesign wurde eine randomisierte, kontrollierte Studie mit der Einteilung in eine Therapie- und eine Warte-Kontrollgruppe (vgl. Abbildung 6) gewählt, da gemäß der zu Grunde liegenden Logik von Leitlinien nur in der EBM-Hierarchie weit oben angesiedelte Studiendesigns in der Formulierung von Therapieempfehlungen Berücksichtigung finden.

Da die im Regelfall aufgeführten 50 Therapieeinheiten mit einer Frequenz von 1-2 Sitzungen pro Woche zu einer maximalen Therapiedauer von einem Jahr führen, wurden 12 Monate als Beobachtungszeitraum gewählt. Dabei soll die Hälfte der eingeschlossenen stotternden Kinder zwischen 7 und 11 Jahren sofort mit der Therapie beginnen, während die andere, zufällig ausgewählte Hälfte 3 Monate auf den Therapiestart wartet. Im weiteren Verlauf werden alle Probanden gleichbehandelt, indem jeweils 3, 6 und 12 Monate nach Therapiebeginn Studienmessungen erfolgen. Das Therapieende ist in diesem Studiendesign nicht festgelegt und kann innerhalb oder außerhalb des Beobachtungszeitraums liegen.

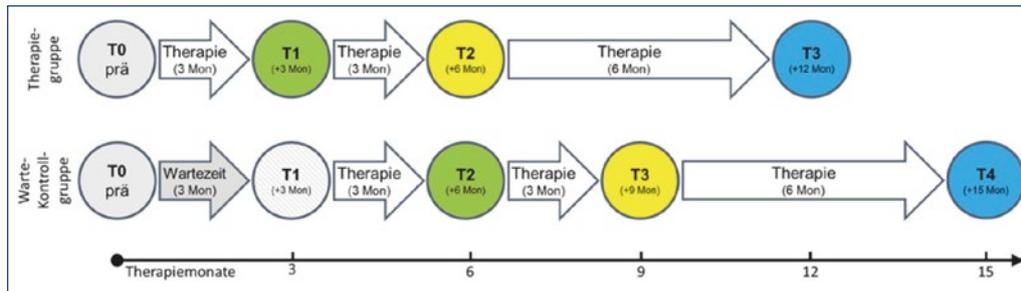


Abb. 6: Studiendesign von PMS KIDS

Die Beurteilung des Therapiefortschritts erfolgt in PMS KIDS ausschließlich quantitativ. Als primäres Outcome wird die psycho-soziale Belastung über einen standardisierten Fragebogen (OASES-S, Euler et al., 2016) gemessen, während als sekundäres Outcome die Stottersymptomatik (SSI-4, Riley, 2009; Schweregradskalen) in verschiedenen Sprechsituationen erfasst wird.

4.3 Einordnung von PMS KIDS in die Versorgungsforschung

PMS KIDS stand von Beginn an vor der Herausforderung, Therapieverläufe unter Alltagsbedingungen zu erfassen und gleichzeitig die notwendigen Anforderungen zu erfüllen, die in randomisierten, kontrollierten Studien an die interne Validität gestellt werden.

Wie methodisch üblich wurde vorab eine Fallzahlkalkulation durchgeführt, um die erforderliche Stichprobengröße bestimmen zu können. Die daraus resultierende Anzahl von 52 Kindern musste um einen Sicherheitspuffer ergänzt werden, da aus anderen Studien bekannt ist, dass gerade bei längeren Studien die sogenannten Drop-out-Zahlen um die 20 % liegen können. Eine vorab durchgeführte Befragung (Kohmächer, 2017) deutete bereits darauf hin, dass einerseits in sprachtherapeutischen Einrichtungen generell wenige stotternde Kinder zeitgleich behandelt werden und nur etwa alle 4 Monate Neuanmeldungen von Kindern dieser Altersgruppe erfolgen. Dementsprechend wurde ein 14monatiger Rekrutierungszeitraum eingeplant, der schlussendlich um 4 Monate verlängert werden musste, um 73 Kinder einschließen zu können. Diesen 73 Kindern stehen 37 Sprachtherapeuten aus dem gesamten Bundesgebiet gegenüber, die sich bereit erklärten, in dem angekündigten Rekrutierungszeitraum alle sich neu anmeldenden Kinder und ihre Eltern über PMS KIDS zu informieren und bei Einwilligung ihre Stottertherapien evaluieren zu lassen. Letztendlich meldeten sich bei 32 dieser sprachtherapeutischen Fachkräfte stotternde Kinder an, was verdeutlicht wie wenig Einfluss das Projektteam trotz Öffentlichkeitsarbeit auf die Anmeldungen in den Praxen hatte.

In der Versorgungsforschung wird gemäß einer systemtheoretischen Sichtweise auf das Versorgungsgeschehen den sogenannten Input-Faktoren, also u.a. den Patienten, Fachkräften, der Organisation und dem System, eine große Bedeutung für das Versorgungsgeschehen und die Wirkung von Gesundheitsinterventionen beigemessen (Gräßel et al., 2015). Dies spiegelt sich in PMS KIDS in einer ausgeprägten Heterogenität der Teilnehmenden wider. Zunächst wurde über breite Ein- und Ausschlusskriterien versucht, möglichst allen sich anmeldenden Kindern und ihren Eltern die Teilnahme an der Studie zu ermöglichen und damit das tatsächliche Versorgungsgeschehen zu erfassen. Dadurch variieren beispielsweise das Alter, das Geschlecht, der kulturelle, sprachliche und sozio-ökonomische Hintergrund, aber auch der Leidensdruck und die Ausprägung der sicht- und hörbaren Stottersymptomatik bei den eingeschlossenen Kindern. Auch die Eltern und andere Familienangehörige spielen für den Therapieerfolg eine wesentliche Rolle, da sie auf unterschiedliche Art und Weise die Therapie, z. B. über die Durchführung von Hausaufgaben, unterstützen und unterschiedlich mit dem Stottern ihres Kindes umgehen.

Die hohe Anzahl beteiligter sprachtherapeutischer Fachkräfte weist darauf hin, dass jeder Therapieverlauf als Einzelfall betrachtet werden muss, der von der Beziehung zwischen therapeutischer Fachperson und Kind in einem bestimmten Kontext geprägt ist. Auch auf Seite der Fachpersonen wurden die Einschlusskriterien breit angelegt, um das tatsächliche Versorgungsgeschehen zu erfassen. Dementsprechend ist die Expertise dieser sprachtherapeutischen Fachkräfte genau wie ihr Alter und Geschlecht sehr heterogen und reicht von wenig Therapieerfahrung (mit Stottern) bis zur ausschließlichen Spezialisierung auf Stottertherapien.

Einen wesentlichen, im Kontext klinisch-experimenteller Studien wichtigsten, Wirkfaktor stellt das eigentliche Therapieverfahren, in diesem Fall *KIDS* dar. Wie viele andere sprachtherapeutische Konzepte gibt *KIDS* einen Rahmen vor und beschreibt Inhalte sowie Phasen, fordert

jedoch von sprachtherapeutischen Fachkräften eine individualisierte Anpassung des Therapieverfahrens an die Bedürfnisse des Kindes (Sandrieser & Schneider, 2015). Eine derartige, in der Versorgungsforschung als komplexe Intervention in einem komplexen Kontext bezeichnete Vorgehensweise kann und sollte bei den 73 stotternden Kindern nicht komplett identisch sein. Dennoch fordert die methodische Strenge eines RCTs eine beschreibbare und vergleichbare Intervention, um die interne Validität der Ergebnisse sicherzustellen. Um dieser Anforderung zu genügen, wurde ein 200seitiges Therapiemanual mit dazugehörigen Therapiematerialien entwickelt und in einer Schulung den Beteiligten nähergebracht. Das therapeutische Fachpersonal wurde instruiert, weiterhin ihre Therapien so zu gestalten, wie sie es gewohnt sind, sich dabei aber anhand des Manuals rückzuversichern, ob ihr Vorgehen dem Ansatz von KIDS entspricht.

4.4 Verlauf von PMS KIDS

Die Einordnung von PMS KIDS als Versorgungsstudie zeigte bereits, dass die Wirkung einer Intervention, in diesem Fall der Stottermodifikationstherapie KIDS, von vielfältigen Input-Faktoren abhängt und nicht losgelöst vom Kontext der Versorgungssituation betrachtet werden kann. Die explizite Berücksichtigung dieser Versorgungssituation macht die Versorgungsforschung aus und sorgt dafür, dass Ergebnisse aus solchen Studien unmittelbar auf die Versorgungspraxis übertragbar sind (vgl. Kapitel 1).

Die Erfahrungen mit PMS KIDS als einer noch andauernden Versorgungsstudie bestätigen, dass sich im Zeitraum der Datenerhebung immer wieder Veränderungen ergeben, die die eigentliche Versorgungsleistung (= die konkreten Therapieverläufe) beeinflussen. So kam es beispielsweise zum Ausfall von drei therapeutischen Fachkräften aufgrund von Krankheit bzw. Schwangerschaft sowie zu neuen Indikationen im Therapieverlauf (z. B. Poltern, Wortfindungsstörung). Wie zuvor erwartet, entschieden sich einige Teilnehmende (n= 16) im Verlauf gegen eine weitere Studienteilnahme (z. B. kein Therapiebedarf, Unzufriedenheit mit der Lösung in die Wartegruppe) bzw. wurden Teilnehmende durch das Projektteam, z. B. aufgrund von mangelnden Deutschkenntnissen, ausgeschlossen.

Über bestimmte Maßnahmen wurde und wird versucht, die individuellen Therapieverläufe zu dokumentieren, um neben der Endwirkung auch einige Angaben über den Prozess machen zu können. So füllen die therapeutischen Fachkräfte nach jeder Therapieeinheit einen kurzen Ankreuzbogen zu den wesentlichen Therapieelementen der Stunde aus. Dies erhöht seitens des Projektteams die Sicherheit, dass Therapien tatsächlich nach KIDS durchgeführt wurden, und erlaubt später eine (grobe) Auswertung der Therapieverläufe. Da nur in den ersten 3 Monaten ein Vergleich zwischen einer Therapie- und einer Wartegruppe erfolgt, wurden die Fachpersonen gebeten, die von ihnen aufgegebenen Hausaufgaben und die reale Umsetzung ihrer Kinder zu dokumentieren. Allerdings zeigt sich hier, ähnlich wie in der Dortmunder Mutismusstudie, dass diese Aufzeichnungen unvollständig und zum Teil gar nicht erfolgten.

Die Corona-Pandemie und die damit einhergehenden Beschränkungen haben sich mit Sicherheit am gravierendsten auf PMS KIDS ausgewirkt und demonstrieren, dass in Studien unter Alltagsbedingungen erhebliche Einflüsse wirken können. Zunächst sind lockdownbedingt Therapien ausgefallen und/oder wurden später durch Videotherapien ergänzt bzw. ersetzt. Gerade bei der komplexen Redeflussstörung Stottern haben die Kontaktbeschränkungen vermutlich zu sehr individuellen Auswirkungen bei den beteiligten Kindern geführt: einige waren durch den Wegfall sprachlicher Anforderungen entlastet, andere fühlten sich durch die Situation insgesamt belastet. In Videotherapien kann der in *KIDS* wesentliche Transfer von Erlerntem in Sprechsituation außerhalb der Therapie nur eingeschränkt angeleitet und geübt werden. Es bleibt abzuwarten, wie sich dies auf die Effekte auswirkt.

Bereits jetzt ist dem Projektteam klar geworden, dass die Komplexität einer Versorgungsstudie nicht vorhersehbar, die Methodik bei einer solchen Studie jedoch im Verlauf nicht beliebig anpassbar ist. Im Nachhinein wäre es wünschenswert gewesen, qualitative Methoden in der Datensammlung zu ergänzen und den Outcome zusätzlich aus Sicht der sprachtherapeutischen Fachkräfte bewerten zu lassen. Eine genauere Prozessevaluation war in dieser Studie nicht möglich, wäre aber eine Bereicherung in der Interpretation der Wirkung in Bezug auf den möglichen Einfluss des Kontextes (Farin et al., 2017).

Zusammenfassend zeigt dieses Beispiel, dass Versorgungsstudien eine sinnvolle Ergänzung zu Einzelfallstudien darstellen, da sie eine starke Einzelfallorientierung aufweisen. Wie auch Einzelfallstudien weisen sie eine große Nähe zur realen Versorgungssituation auf. Der Vorteil von Gruppenstudien wie PMS KIDS liegt darin, dass die Ergebnisse besser zu verallgemeinern sind

als Einzelfallstudien. Demgegenüber steht der Nachteil, dass die Datenerhebungsmethoden aufgrund der Stichprobengröße beschränkt werden müssen und der Therapieprozess weniger detailliert evaluiert werden kann. Zuletzt sei auf die Schwierigkeit hingewiesen, die Effekte aus Versorgungsstudien wie PMS KIDS mit klinisch-experimentellen Studien zu vergleichen. Durch den bewusst nicht kontrollierten Einfluss des Kontextes in Versorgungsstudien fallen die Effekte häufig niedriger aus als in klinisch-experimenteller Forschung (effectiveness gap). Im Englischen wird dieser Unterschied sehr treffend in den beiden Begriffen *efficacy*, Wirksamkeit im kontrollierten klinischen Versuch, und *effectiveness*, Wirksamkeit unter Alltagsbedingungen, ausgedrückt.

5 Evaluation von Redefluss- und StotterKompass als Beispiel für eine kumulierte Einzelfallstudie im mixed-method-Design

Bevor das Studiendesign beschrieben und als kumulierte Einzelfallstudie vorgestellt wird, sollen die Kompass und deren Einsatz kurz dargestellt werden.

Redefluss- und StotterKompass sind fragebogenähnliche Instrumente, die im komplementären Einsatz zur Früherkennung, Ersterfassung und Erstberatung bei beginnendem Stottern benutzt werden können. Der RedeflussKompass (RfK) ist für ein Laienzielpublikum konzipiert und kann von Bezugspersonen (= Eltern, pädagogische Fachkräfte etc.) benutzt werden. Er hat die Funktion eines Risikoscreenings und beantwortet durch das quantifizierte Ergebnis die Frage, ob es sich um beginnendes Stottern handelt. Der komplementäre Einsatz beider Kompass ist so gedacht, dass die Bezugspersonen zunächst alleine den RedeflussKompass bearbeiten und mit dem Ergebnis des RfK zum Erstgespräch mit der sprachtherapeutischen Fachkraft gehen. Der ausführlichere StotterKompass (SK) ist für die Fachperson konzipiert und dient als Gesprächsleitfaden für das Erstgespräch. Er hat das Potential, die Ergebnisse des RedeflussKompasses zu validieren, da er alle Items des RfK enthält und ebenfalls ein quantifiziertes Ergebnis generiert. Der StotterKompass beansprucht, eine Hilfe für die Ersterfassung und Erstberatung zu sein (für eine ausführlichere Darstellung siehe Kohler & Braun 2019).

5.1 Forschungsdesign der Studie

Ziel der Studie war es, Gütekriterien für die Instrumente zu definieren und den Instrumenteneinsatz kritisch-konstruktiv zu reflektieren (Kohler & Braun, 2020).

Auch wenn die Studie nicht so recht in die Evidenzhierarchie (vgl. Abbildung 1) eingeordnet werden kann, weil sie nicht die Wirksamkeit einer Methode zum Gegenstand hat, gehört sie zur Evidenzforschung, weil sie die Güte von Entscheidungen zum Thema macht. Die Überprüfung der Kompass geschah mit $N = 60$ Einzelfällen und hatte den Charakter eines quasi-experimentellen Settings. Die Durchführung der Kompass stellte die unabhängige Variable dar und die quantifizierten Ergebnisse die abhängige Variable. Die Ergebnisse wurden mit dem ebenfalls quantifizierten wahren Wert des tatsächlichen Diagnostik-, Beratungs- bzw. Therapiegeschehens konfrontiert, so dass mit Hilfe von statistischen Kennwerten die Gütekriterien Sensitivität, Spezifität sowie Kriteriumsvalidität bestimmt werden konnten. Ausserdem konnte die Inter-Rater-Reliabilität für den StotterKompass berechnet werden, da bei 17 Fällen eine voneinander unabhängige Zweidurchführung des Instrumenteneinsatzes möglich war.

Neben diesen Strategien der quantitativen Sozialforschung wurden auch qualitative Strategien eingesetzt. Es wurden die Erfahrung der Anwender (Bezugspersonen mit dem RfK und Fachpersonen mit dem SK) erfragt und zum Teil auch beobachtet. Abbildung 7 stellt das Forschungsdesign dar. Die Pfeile, welche von dem grün unterlegtem Kästchen der Qualitativen Dokumentenanalyse ausgehen, sollen aussagen, dass mit den qualitativen Ergebnissen Erklärungen für das Zustandekommen der quantitativen Ergebnisse möglich waren. In diesem Sinne hatte die qualitative Strategie eine zeitlich nachgeordnete Funktion. Grundsätzlich waren quantitative und qualitative Strategien aber zeitlich parallel angelegt und gleichwertig in der Gewichtung des gesamten Mixed-Methods-Forschungsprozesses (Kuckartz, 2014).

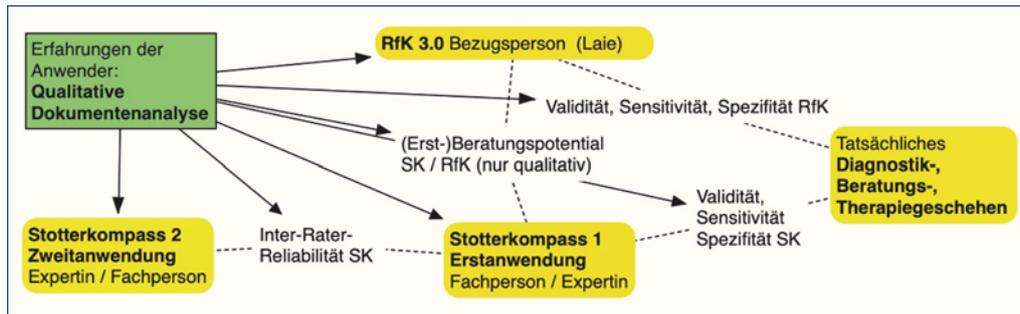


Abb. 7: Forschungsdesign der Evaluation der Kompass (vgl. Kohler & Braun, 2019)

Im Folgenden werden einige ausgewählte methodische Aspekte und Ergebnisse des Forschungsprojekts unter Rückgriff der in Kapitel 1 gemachten Ausführungen dargestellt und diskutiert. Der Schwerpunkt liegt dabei auf der kritischen Analyse des Zustandekommen der qualitativen Kategorien in Gegenüberstellung zu ihrer Nützlichkeit.

5.2 Dokumentenanalyse – Zustandekommen der qualitativen Kategorien

Die qualitativen Erhebungsstrategien waren äusserst vielfältig und orientierten sich an dem Prinzip des nice-to-have. Es war z. B. nicht allen Fachpersonen zu jedem Fall möglich, ein Interview zu geben. Insgesamt standen 35 Protokolle oder Audioaufnahmen von Interviews mit den Fachpersonen zur Verfügung. Ebenso kam auch nicht mit allen Bezugspersonen ein Interview zustande. Für die Auswertung lagen letztendlich 17 Dokumentationen von Interviews mit Bezugspersonen vor. Die Dokumentationen hatten eine unterschiedliche Güte. Mal waren es Protokolle oder Audioaufnahmen wie bei den Fachpersonen, in einigen Fällen lagen auch methodisch hochwertige Transkriptionen vor. Bei den Fachpersonen gab es neben den Dokumentationen der Interviews neun relativ ausführliche schriftliche Reflexionen, welche die Erfahrungen bei der Durchführung des Stotterkompass betrafen. Man sieht an der Zahl der Dokumente pro Teilstichprobe, dass die Perspektive der Bezugspersonen weniger stark vertreten war im Vergleich zur Perspektive der Fachpersonen. In der Perspektive der Fachpersonen bildete sich aber häufig die Perspektive der Bezugspersonen indirekt ab, indem die Fachpersonen von den Rückmeldungen der Bezugspersonen berichteten. Bei der Kategorienbildung wurden beide Perspektiven meist zusammengenommen und eine Mindesthäufigkeit von zehn Nennungen festgelegt, um überhaupt in das Kategoriensystem aufgenommen zu werden.

Manche Kategorien hatten also größere Chancen, genannt zu werden als andere, insofern ist es diskussionswürdig, die Häufigkeiten überhaupt miteinander zu vergleichen. Die kritische Zahl von 10 als eine ordinale Grösse ist willkürlich und kann inhaltlich nur insofern begründet werden, als die Forschenden damit eine übermässige Kategorienanhäufung verhindern wollten und in den unterschiedlichen Häufigkeiten trotz aller Unwägbarkeiten beim Zustandekommen eine unterschiedliche Gewichtung der Kategorien deutlich machen wollten.

Unter dem Gesichtspunkt der methodischen Strenge (Döring & Bortz, 2016) ist diese Art von Datenaufbereitung und Auswertung also kritisierbar. Legt man das Kriterium der Nützlichkeit einer anwendungsorientierten Mixed-Method-Strategie an (Döring & Bortz, 2016), sieht es aber anders aus. In der gewichteten Abwägung zwischen methodischer Strenge und Nützlichkeit hatten sich die Autoren für letzteres entschieden. Im Folgenden werden daher drei ausgewählte qualitative Kategorien bzw. Ergebnisse in ihrer Nützlichkeit für die Optimierung und für die Handhabung des Instrumentariums dargestellt. Zum Abschluss wird das Zustandekommen der statistischen Kennwerte als Gütezeichen kritisch diskutiert.

5.3 Nützlichkeit

5.3.1 Erstes Beispiel – Itemdifferenzierung

Zwei der Kategorien betreffen die Randnotizen, welche von den Fachpersonen und von den Bezugspersonen auf den Kompassen gemacht wurden. Die Randnotizen hatten inhaltliche Beziehungen zu anderen Kategorien, da sie meist ein Problem in der Handhabung der Kompass betrafen. Der in diesem Beispiel relevante Inhalt betrifft die Kategorie „zeitliche Dimension“. Mit Abbildung 8 wird eine Randnotiz auf dem Stotterkompass veranschaulicht, welche das Problem der fehlenden Unterscheidung von Qualität und Quantität der Symptomatik auf der Dimension Zeit bei SK und RfK zum Gegenstand hat.

| Veränderungen der Unflüssigkeiten seit dem ersten Auftreten | | | |
|---|---|--------------|---|
| 4.2 Hat sich, seit Sie die auffälligen Unflüssigkeiten beobachten, etwas verändert? | Die Unflüssigkeiten waren zu Beginn häufiger und haben mit der Zeit abgenommen. | 0 | Je stärker die Zunahme der Unflüssigkeiten desto geringer die Wahrscheinlichkeit einer Remission = Eltern haben im RfK nichts gemerkt, da da keine Anhalt geben sollten? → Variation? |
| | Die Häufigkeit der Unflüssigkeiten hat sich seit ihrem Auftreten nicht verändert. | 1 | |
| | Die Unflüssigkeiten waren zu Beginn weniger häufig und haben sich eher abgenommen. | 2 | |
| | Die Unflüssigkeiten und die einhergehende Anstrengung waren zu Beginn weniger und haben mit der Zeit deutlich zugenommen. | 3 | |
| Reaktionen auf die Unflüssigkeiten | | | |
| 4.3 Wie reagiert das Kind auf die Unflüssigkeiten? | Es bemerkt die Unflüssigkeiten gar nicht. | 0 | Je ausgeprägter die negative Reaktionen des Kindes desto höher der Schweregrad des Stotterns fischer: ja, obkehl hat |
| | Es bemerkt die Unflüssigkeiten aber macht nichts Grossartiges daraus. | 1 | |
| | Es bemerkt die Unflüssigkeiten und reagiert emotional (z.B. Angst, Wut, Scham). | 3 | |
| 4.3.1 Das Kind vermeidet aus Angst vor dem Stottern Sprechsituationen. | nie | 0 | Je ausgeprägter das Vermeidverhalten des Kindes desto gravierender sind die Einschränkungen in der sozialen Partizipation und Teilhabe |
| | selten | 1 | |
| | manchmal | 2 | |
| | häufig | 3 | |
| 4.3.2 Das Kind vermeidet aus Angst vor dem Stottern einzelne Wörter | nie | 0 | |
| | selten | 1 | |
| | manchmal | 2 | |
| | häufig | 3 | |
| 4.3.3 Das Kind vermeidet während des Stotterns den Blickkontakt mit dem Gesprächspartner | nie | 0 | |
| | selten | 1 | |
| | manchmal | 2 | |
| | häufig | 3 | |
| 4.3.4 Das Kind verändert wegen des Stotterns seine Sprechweise in auffälliger Art und Weise (Flüstern, langsames Sprechen etc.) | nie | 0 | |
| | selten | 1 | |
| | manchmal | 2 | |
| | häufig | 3 | |

Veränderte Stotterstruktur

Abb. 8: Randnotizen auf dem Stotterkompass

Item 4.3 will die emotionalen Reaktion des Kindes auf die Unflüssigkeit erfassen. Dabei hat die Fachperson eine Randnotiz gemacht, welche die zeitliche Dimension klärt. Diese Art von Randnotiz kam häufig vor und auch in den Rückmeldungen der Bezugspersonen und Fachpersonen wurde das Problem der Abhängigkeit der emotionalen Reaktionen von der Zeit häufig thematisiert. Insgesamt wurde diese Rückmeldung 25 mal gegeben. Dies veranlasste die Autoren der Kompass dazu, in der Folgeversion die Dimension Zeit zu operationalisieren und ein Item Variabilität einzufügen, welches der Situationspezifität und Phasenabhängigkeit des beginnenden Stotterns gerecht wird. Ausserdem wurde die Möglichkeit der Realisierung von Randnotizen durch entsprechende Felder in der weiterentwickelten Versionen der Kompass ausdrücklich ermöglicht.

5.3.2 Zweites Beispiel – Optimierung von Skalierung und Farbgestaltung

Zwei Kategorien betreffen die Skalierung und das Farbsystem der Kompass. Abbildung 9 zeigt beides am Beispiel des Items zur Primärsymptomatik Blockierung.

| | | Skalierung | Farbsystem |
|---|---------------------------------|------------|------------|
| Es scheint so, als ob die Pause unfreiwillig geschieht und das Kind „hängen bleibt“, blockiert. | „Ich k ... kann gut schwimmen.“ | nie | 0 |
| | | selten | 1 |
| | | manchmal | 2 |
| | | häufig | 3 |

Abb. 9: Skalierung und Farbsystem bei RfK und SK

Die Skalierung ist wichtig, um den Schweregrad der Primärsymptomatik zu erfassen, jedoch konnten sich viele Bezugspersonen keinen Unterschied zwischen selten und manchmal vorstellen. Es war also mehr oder weniger Zufall oder von anderen Motiven geleitet, ob sich solche Anwender für einen Punkt (= selten) oder für zwei Punkte (= manchmal) entschieden, aber es gab bei fehlender Differenzierung zwischen diesen beiden Begriffen keinen inhaltlich angemessenen

Grund für die Wahl. Dieses Problem der Skalierung existierte auch bei anderen Items, so dass die Reliabilität gefährdet ist.

Die Farben, welche das ansteigende Gefahrenpotential bzw. den ansteigenden Schweregrad hier visuell verstärken sollten, lösten eine differenzierte Bewertung aus. Einige das Instrument anwendende Personen fanden die Ampelfarben nützlich, die meisten Bezugspersonen allerdings erschrakten bei häufigem rot und fühlten sich unwohl („ich sehe nur noch rot!“).

Diese Rückmeldungen der nutzenden Personen führten dazu, dass die Autoren die begriffshängige Skalierung durch einen Schieberegler ersetzen, der nur noch mit zwei Polen des Schweregrades oder der Intensität bestückt ist. Das Farbsystem wurde bei den Items völlig eliminiert und taucht in der optimierten Version der Kompass nur noch bei der Endauswertung auf.

5.3.3 *Drittes Beispiel – Abwägung unterschiedlicher Durchführungsoptionen*

Dieses Beispiel hat keine Referenz zum Kategoriensystem. Es beruht auf Rückmeldungen der Fachpersonen, die weniger als zehnmals formuliert wurden und doch erschien es den Autoren so wichtig, dass sie daraus eine Konsequenz zogen.

Durch die Interviews mit den Fachpersonen erfuhren die Forschungsleitenden von unterschiedlichen Formen der Durchführung. Eine Form gehorchte einem sequentiellen Vorgehen. Die Fachpersonen hielten sich bei ihrer Gesprächsführung an die Ordnung der Items im StotterKompass und arbeiteten diese nacheinander (sequentiell) ab. Andere Fachpersonen gestalteten die Gesprächsführung flexibel. Das heißt, sie benutzten den StotterKompass zwar als Gesprächsleitfaden, aber ohne die Items nacheinander abzuarbeiten.

Beide Formen haben Vor- und Nachteile. Bei der Durchführung in sequentieller Form berichtete eine Fachperson, dass sie weniger in der Lage war, auf die Wünsche der Bezugspersonen einzugehen und sie diese Form als überstrukturiert für ein Erstgespräch betrachte. Andererseits gibt die sequentielle Form eine Klarheit im Ablauf des Gesprächs. Es ist weniger wahrscheinlich, den Faden der Gesprächsführung zu verlieren oder ein wichtiges Item zu vergessen. Das wäre ein möglicher Nachteil der flexiblen Gesprächsführung, die ihrerseits als Vorteil mehr Raum für das empathische Zuhören mitbringt.

Diese Abwägung der Vor- und Nachteile, welche von den Fachpersonen geäußert wurden, war den Autoren so wichtig, dass sie beispielhaft in die Handanweisung der Weiterentwicklung des StotterKompasses aufgenommen wurde.

5.4 Zwischenfazit

Alle drei Beispiele zeigen, wie qualitative Daten zur Weiterentwicklung und Optimierung des Instrumenteneinsatzes beitragen können. Sie sind damit nicht nur nützlich, sondern erhöhen die Validität und Reliabilität der Instrumente. Dieser Nutzen von qualitativen Strategien gilt nicht nur für die Evidenz von Screening- oder Diagnostikinstrumenten, sondern könnte auf die Wirksamkeitsforschung von Intervention übertragen werden. Das letzte Beispiel der unterschiedlichen Durchführungsoptionen erscheint für andere Interventionsformen ebenfalls relevant, denn auch bei sprachtherapeutischen Interventionen kann im Einzelfall eine Spannung zwischen den Vorgaben eines Therapiemodells und einer notwendigen Flexibilisierung des Vorgehens existieren.

Trotzdem bleibt das schon oben angeführte Dilemma zwischen eingeschränkter methodischer Strenge und hoher Nützlichkeit bei der Beurteilung dieser kumulierten Einzelfallstudie. Dies könnte eine zentrale Diskussion bei der Verwendung von kumulierten Einzelfallstudien im Mixed-Method-Design für die Sprachtherapie sein. Dabei geht es nicht nur um qualitative Daten und deren Erhebung und Auswertung, sondern auch um quantitative Strategien und deren Wert. Als letztes Beispiel sei daher ein Problem der methodischen Strenge im Zusammenhang mit der Operationalisierung des wahren Wertes für die Berechnung der Gütekriterien angeführt.

5.5 Wie wahr ist der wahre Wert?

Der „wahre“ Wert ist fundamental für das Zustandekommen von Sensitivität, Spezifität und Kriteriumsvalidität (Kohler, 2017). Bei dem hier vorgestellten Forschungsprojekt handelt es sich um den tatsächlichen Diagnostik, Beratungs- bzw. Therapiebedarf. Um diesen festzustellen, ist eine ausführliche sprachtherapeutische Diagnostik der Goldstandard, die idealerweise völlig unabhängig vom zu evaluierenden Instrument durchgeführt wird. Die Verwirklichung des Goldstandards war jedoch nicht möglich. Insbesondere wenn bei dem Screening mit dem RfK und dem Erstgespräch unter Zuhilfenahme des SK kein weiterer Handlungsbedarf ermittelt wurde, waren die Bezugspersonen nicht immer bereit, sich auch noch einer differenzierten Diagnostik zu un-

terziehen. Wir hatten daher ein Befragungsprozedere mit den involvierten Fachleuten entwickelt, um diesen tatsächlichen Bedarf mit höchstmöglicher Sicherheit zu erfassen. Wir nutzten zudem auch Videoaufnahmen und wo immer möglich, die Beobachtung, um die Güte des wahren Wertes so maximal wie möglich zu gestalten. Trotzdem ist dieses Vorgehen unterhalb der Güte des idealen Goldstandard angesiedelt. Es gelang auch nicht in jedem Fall eine genügende Sicherheit für die Gültigkeit des wahren Wertes zu erlangen, so dass einige Fälle ausgeschlossen werden mussten. Die Diskussion zur Operationalisierung des wahren Wertes ist damit wie bei der Erhebung von qualitativen Daten und deren Auswertung, geprägt von dem Ideal einer methodischen Strenge und der Machbarkeit im Rahmen eines Forschungsprojektes.

6 Zusammenfassung und Diskussion

Die in diesem Artikel aufgeführten Studien verdeutlichen die Bedeutung der Einzelfallorientierung in der Sprachtherapie. Sprachdiagnostiken und -therapien finden immer in einem (schulischen oder therapeutischen) Kontext statt, der die Ausgestaltung maßgeblich mit beeinflusst. In der Dortmunder Mutismusstudie (Kapitel 3) wurde herausgestellt, dass sich beim Mutismus kommunikative Veränderungen in Kontexten unterschiedlich ausgeprägt darstellen, so dass unterschiedlichste Situationen und auch unterschiedliche Qualitäten des sprachlich-kommunikativen Verhaltens dokumentiert werden müssten, um den Therapieerfolg zu erfassen. Die PMS KIDS-Studie (Kapitel 4) stand vor einer ähnlichen Herausforderung, da eine Entscheidung getroffen werden musste, in welchen Situationen (z. B. im Therapiesetting oder zu Hause) das Sprechverhalten erfasst werden soll. Auch bei der Beurteilung einer Stottersymptomatik anhand der Redefluss- und StotterKompass muss das zeit- und situationsspezifische Verhalten mitberücksichtigt werden (Kapitel 5). Es zeigt sich damit, dass sprachtherapeutische Störungsbilder, wie hier am Beispiel Redefluss- und Kommunikationsstörungen, komplexer Natur sind und damit eine komplexe Diagnostik sowie sprachtherapeutisches Vorgehen erfordern. Sprachtherapien können also nur dann erfolgreich sein, wenn sie die spezifische Problematik und die Wünsche der zu behandelnden Person adäquat berücksichtigen (soziale Evidenz, vgl. Beushausen & Grötzbach, 2018) und ihre Therapien immer wieder an neue Gegebenheiten anpassen. Dies erfordert von Sprachtherapeuten eine hohe Expertise, auch interne Evidenz genannt (Beushausen & Grötzbach, 2018).

Ein derartig einzelfallorientiertes Vorgehen lässt sich mit Einzelfallstudien sehr gut evaluieren. Ausführliche Datenerhebungen sind genauso wie ein Mix aus quantitativen und qualitativen Erhebungsmethoden möglich. Zusätzlich können Sprachtherapeuten Rückmeldungen über ihr Vorgehen geben, wodurch eine zusätzliche Evaluation des Prozesses möglich wird (Beushausen, 2014). Diese systematisierten Angaben haben das zusätzliche Potential zur Optimierung von Therapieansätzen, -modellen oder -programmen, indem notwendige Anteile der Intervention ausgemacht und gleichzeitig alternative Vorgehensweisen entwickelt werden. In der Versorgungsforschung werden solche Prozessevaluationen standardmäßig neben Wirksamkeitsstudien gefordert, um den Implementierungsprozess ggf. anpassen zu können (Farin et al., 2017). Auf diese Weise käme der sprachtherapeutischen Forschung auch eine Funktion der Qualitätssicherung und -weiterentwicklung zu.

Der Nachteil der fehlenden Verallgemeinerbarkeit von Einzelfallstudien kann zumindest teilweise über die Kumulation mehrerer Einzelfälle ausgeglichen werden (vgl. Kap. 2.4). Kumulative Versorgungsstudien wie PMS KIDS erlauben ebenfalls verallgemeinerbare Aussagen und tragen der Komplexität der Kontexte und Interventionen gleichzeitig explizit Rechnung. Dem Vorteil der Verallgemeinerbarkeit steht dann allerdings gegenüber, dass die Einzelfälle und -verläufe weniger ausführlich dokumentiert und gemessen werden können. Auch wird die Handhabung eines Manuals erforderlich, um die interne Validität sicherzustellen. Hier lässt sich diskutieren, ob die Nutzung eines solchen Manuals, selbst wenn es keine strikten Therapievorgaben macht, nicht zu einer Entfernung der therapeutischen Fachpersonen von ihrem eigentlichen, alltäglichen Vorgehen führt.

Aus Sicht der Autoren ist die Orientierung am Einzelfall in der Evidenzbasierung der Sprachtherapie eine Pflicht. Um Evidenzen zu generieren, haben unterschiedliche Studiendesigns (in unterschiedlichen Forschungsphasen) ihre Berechtigung (Beushausen, 2014) und sollten sich Forschende nicht von der Evidenzhierarchie dazu verführen lassen, Einzelfallstudien als wenig intern valide auszuklammern. Vielmehr sollte die Fragestellung die Wahl des Studiendesigns

bestimmen und eine Abwägung zwischen den Anforderungen an interne und externe Validität getroffen werden. Döring und Bortz (2016) formulieren dies wie folgt:

„Eine faire Bewertung bemisst Studien also nicht pauschal an Methoden-Idealen, sondern vor dem Hintergrund ihrer selbst gesetzten Erkenntnisziele und unter Beachtung der forschungsökonomischen Rahmenbedingungen (etwa Zeit und Kostenplan). Bevor man eine Studie wegen deren methodischen Einschränkungen und Schwächen grundsätzlich verwirft, kann es aufschlussreich sein, sich die Frage zu stellen „Was ist/war die methodische Alternative?“ Limitationen sind dann zu akzeptieren, wenn eine strengere Methodik aus organisatorischen oder ethischen Gründen nicht realisierbar war und wenn methodische Grenzen bei der Ergebnisinterpretation ausdrücklich einbezogen werden.“ (Döring & Bortz, 2016), S. 22)

Neben den forschungsökonomischen Abwägungen stellt sich auch im Verlauf von (sprachtherapeutischen) Studien die Frage, inwiefern vom geplanten Vorgehen abgewichen werden darf, um veränderten Kontextbedingungen Rechnung tragen zu können. Während dies inhaltlich sehr sinnvoll sein kann und in qualitativen Studien üblich ist, können ethische Überlegungen (wie viel Mehraufwand ist den Probanden zuzumuten?) und die Registrierung einer Studie, z. B. im Deutschen Register Klinischer Studien, ein solches Vorgehen verhindern. Hier zeigt sich, dass eine stärkere Gewichtung des qualitativ-zirkulären Vorgehens eine gute Ergänzung zum eher quantitativ-linearen Forschungsprozess darstellen könnte. Der Forschungsprozess würde sich dem therapeutischen Prozess in seiner Komplexität mehr annähern und könnte ihn in seiner Realität besser abbilden.

Für die Evidenzbasierung der Sprachtherapie erscheint den Autoren dieses Artikels die Nähe zum tatsächlichen Versorgungsgeschehen, wie sie in Einzelfall- und Versorgungsstudien abgebildet werden kann, als wesentlich. Zu beachten ist in der Interpretation solcher Studienergebnisse jedoch der bereits erwähnte effectiveness gap, also der Unterschied in der Wirkung unter kontrollierten Bedingungen und im Alltag. Hier stellt sich die Frage, ob ein Vergleich von Studien und ihrer Effektgrößen sinnvoll ist, oder ob nicht andere bzw. eigene Richtlinien zur Bewertung der Evidenz herangezogen werden müssten. Vielleicht sollte man die in den einschlägigen Werken zur Forschungsmethodik immer wieder formulierten Warnungen, die statistischen Kennwerte nur unter äußerster Vorsicht miteinander zu vergleichen (z. B. Bühner & Ziegler, 2009; Döring & Bortz, 2016; Eid et al., 2010) so ernst nehmen, dass man auf einen Bezug zu einer Richtlinie ganz verzichtet und stattdessen den gefunden Wert im Kontext der eigenen bzw. nur wirklich vergleichbaren Studien bewertet.

Das Zustandekommen der statistischen Kennwerte wurde auch im Zusammenhang mit der Evaluation der Kompassse kritisch reflektiert. Hier ging es um die Operationalisierung des wahren Wertes, der massgeblich für die Berechnung von Sensitivität, Spezifität und Kriteriumsvalidität ist. Die kritische Frage dazu lautet: Wie weit darf man vom Goldstandard abweichen? Im Rahmen der Einzelfallorientierung könnten hier weitere Gütekriterien in den Vordergrund treten wie zum Beispiel die Inter-Rater-Reliabilität als ein Indikator für die Objektivität. Dieses berechenbare Maß könnte für die einzelfallorientierte Forschung besonders bedeutsam sein, um die fehlenden Kontrollmöglichkeiten der vielfältigen Variablen und damit geringere interne Validität von Praxisforschung zumindest teilweise zu kompensieren. Die Inter-Rater-Reliabilität kann man dabei nicht nur als eine Schnittstelle zwischen den in der quantitativen Sozialforschung wichtigen Güteigenschaften Objektivität und Reliabilität betrachten, sondern auch als eine Schnittstelle zwischen quantitativer und qualitativer Sozialforschung. Die Übereinstimmung zwischen zwei Urteilen wird in der qualitativen Forschung als ein zentrales Gütemerkmal betrachtet. Es betrifft solche Gütekonstrukte wie die kommunikative Validierung (als Akt der Abgleichung von Daten zwischen forschender und der untersuchten Person), die intersubjektive Nachvollziehbarkeit (als eine Form von Abgleichung der Datenlage zwischen verschiedenen, also mindestens zwei unterschiedlichen Rezipienten der Daten) oder der Investigator-Triangulation (als die Forderung nach personeller Perspektivenvielfalt bei der Erhebung, Auswertung und Interpretation von Daten).

Für die Professionalisierung der Therapieberufe ist der Ausbau von Forschung und damit einhergehend der Evidenzbasierung entscheidend. Der Wissenschaftsrat formuliert dazu: „Mit Blick auf die Weiterentwicklung der Forschung unterstützt der Wissenschaftsrat die Empfehlungen des Gesundheitsforschungsrates, der insbesondere in Hinsicht auf die Bereiche der Klinischen Forschung und der Versorgungsforschung Potenziale für genuine, von der Forschung in der Me-

dizin und weiteren Bezugsdisziplinen hinreichend klar abgrenzbare Forschung in den Gesundheitsberufen erkannt hat.“ (2012), S. 98). Es gilt also, eine für die Sprachtherapie eigene Form der Forschung (weiter) zu entwickeln.

Literatur

- Anstendig, K. (1998). Selective mutism: A review of the treatment literature by modality from 1980–1996. *Psychotherapy: Theory, Research, Practice, Training*, 35(3), 381–391. <https://doi.org/10.1037/h0087851>
- Bergman, R. L., Gonzalez, A., Piacentini, J. & Keller, M. L. (2013). Integrated Behavior Therapy for Selective Mutism: A randomized controlled pilot study. *Behaviour Research and Therapy*, 51(10), 680–689. <https://doi.org/10.1016/j.brat.2013.07.003>
- Beushausen, U. (2014). Chancen und Risiken einer evidenzbasierten Praxis. *LOGOS*, 22(2), 96–104.
- Beushausen, U. & Grötzbach, H. (2018). Evidenzbasierte Sprachtherapie (2. überarbeitete Auflage). Idstein: Schulz-Kirchner.
- Borgetto, B., Spitzer, L. & Pfüngsten, A. (2016). Die Forschungspyramide: Evidenz für die logopädische Praxis brauchbar machen. *Forum Logopädie*, 30(1), 24–28.
- Bühner, M. & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson.
- Cohan, S. L., Chavira, D. A. & Stein, M. B. (2006). Practitioner review: Psychosocial interventions for children with selective mutism: a critical evaluation of the literature from 1990–2005. *Journal of child psychology and psychiatry, and allied disciplines*, 47(11), 1085–1097. <https://doi.org/10.1111/j.1469-7610.2006.01662.x>
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. vollständig überarbeitete, aktualisierte und erweiterte Auflage). Berlin: Springer.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden*. Weinheim, Basel: Beltz.
- Euler, H. A., Kohmäscher, A., Cook, S., Metten, C. & Miele, K. (2016). *Overall Assessment of the Speaker's Experience of Stuttering (OASES-S): Protokollbogen*. (deutsche Übersetzung). McKinney (TX): Stuttering Therapy Resources.
- Farin, E., Möhler, R. & Meyer, G. (2017). Doppelte Komplexität: komplexe Interventionen in komplexen Kontexten. In H. Pfaff, E. Neugebauer, G. Glaeske & M. Schrappe (Hg.), *Lehrbuch Versorgungsforschung: Systematik – Methodik – Anwendung* (2. Aufl., S. 84–88). Stuttgart: Schattauer.
- Flick, U., Kardorff, E. von, Krupp, H., Rosenstiel, L. & Wolff, S. (Hg.). (2012). *Handbuch qualitative Sozialforschung: Grundlagen, Konzepte, Methoden und Anwendungen* (3., neu ausgestattete Aufl.). Weinheim: Beltz.
- Gensthaler, A., Dieter, J., Raisig, S., Hartmann, B., Ligges, M., Kaess, M., Freitag, C. M. & Schwenck, C. (2020). Evaluation of a Novel Parent-Rated Scale for Selective Mutism. *Assessment*, 27(5), 1007–1015. <https://doi.org/10.1177/1073191118787328>
- Gräbel, E., Donath, C., Hollederer, A., Drexler, H., Kornhuber, J., Zobel, A. & Kolominsky-Rabas, P. (2015). Versorgungsforschung – evidenzbasiert: Ein Kurzüberblick und Implikationen. *Gesundheitswesen*, 77, 193–199.
- Jain, A. & Spieß, R. (2012). Versuchspläne der experimentellen Einzelfallforschung. *Empirische Sonderpädagogik*, 4(3/4), 211–245.
- Julius, H., Schlosser, R. W. & Goetze, H. (2000). *Kontrollierte Einzelfallstudien: Eine Alternative für die sonderpädagogische und klinische Forschung*. Göttingen: Hogrefe.
- Kasten, H. (Hg.). (2014). *Entwicklungspsychologie: Lehrbuch für pädagogische Fachkräfte* (1. Aufl.). Haan-Gruiten: Verl. Europa-Lehrmittel Nourney.
- Kohler, J. (2017). Evidenz bei der Identifikation des beginnenden Stotterns. *Praxis Sprache*, 3, 147–154.
- Kohler, J. & Braun, W. (2019). Einsatz von RedeflussKompass 3.0 zur Früherkennung und StotterKompass zur Ersterfassung/Erstberatung. *Forschung Sprache*, 3, 3–24.
- Kohler, J. & Braun, W. (2020). *Früherkennung, Ersterfassung und Erstberatung bei beginnendem Stottern. Die Erforschung und Entwicklung von RedeflussKompass und StotterKompass zu einem digitalen Instrumentarium*. Idstein: Schulz-Kirchner.
- Kohmäscher, A. (2017). Wissen was wirkt: Herausforderungen bei der Evaluation von Stottertherapien. In Interdisziplinäre Vereinigung von Stottertherapeuten (Hg.), *Wirksamkeit von Stottertherapien: Was wir tun und wie wir messen* (S. 9–22). Köln: Demosthenes.
- Kohmäscher, A. (18. Mai 2019). *Stottertherapie in ambulanten Einrichtungen – ein Abgleich der aktuellen Versorgungssituation mit den Leitlinien*. dbl-Kongress, Bielefeld.
- Kraus, E. (2018). Internationale Perspektiven auf die Voraussetzungen einer effizienten Diagnostik und Therapie unter dem Anspruch von Teilhabe und Partizipation. In R. Haring & J. Siegmüller (Hg.), *Evidenzbasierte Praxis in den Gesundheitsberufen: Chancen und Herausforderungen für Forschung und Anwendung* (S. 127–146). Berlin, Heidelberg: Springer.
- Kuckartz, U. (2014). *Mixed methods: Methodologie, Forschungsdesigns und Analyseverfahren*. Berlin, Heidelberg: Springer.
- Kulawiak, P. R. & Wilbert, J. (2018). Komplementäre Studiendesigns zur Evidenzbasierung in der Bildungswissenschaft. In R. Haring & J. Siegmüller (Hg.), *Evidenzbasierte Praxis in den Gesundheitsberufen: Chancen und Herausforderungen für Forschung und Anwendung* (S. 17–31). Berlin, Heidelberg: Springer.
- Lamnek, S. & Krell, C. (2016). *Qualitative Sozialforschung: Mit Online-Materialien* (6., überarbeitete Auflage). Weinheim, Basel: Beltz.
- Mackowiak, K. & Lengning, A. (2010). BAV 3-11. *Das Bochumer Angstverfahren für Kinder im Vorschul- und Grundschulalter*. Göttingen: Hogrefe.
- Melfsen, S. & Walitza, A. (2017). Behandlungsmethoden des selektiven Mutismus. *Sprache – Stimme – Gehör*, 41, 91–97.
- Meyer, T. & Flick, U. (2017). Methoden der qualitativen Forschung. In H. Pfaff, E. Neugebauer, G. Glaeske & M. Schrappe (Hg.), *Lehrbuch Versorgungsforschung: Systematik – Methodik – Anwendung* (2. Aufl., S. 77–83). Stuttgart: Schattauer.
- Natke, U. & Kohmäscher, A. (2020). *Stottern: Wissenschaftliche Erkenntnisse und evidenzbasierte Therapie*. Berlin: Springer.

- Neumann, K., Euler, H. A., Bosshardt, H.G., Cook, S., Sandrieser, PP., Schneider, P., Sommer, M. & Thum, G. (2016). *Pathogenese, Diagnostik und Behandlung von Redeflussstörungen: Evidenz- und konsensbasierte S3-Leitlinie*, AWMF-Registernummer 049-013, Version 1. <http://www.awmf.org/leitlinien/detail/ll/049-013.html>
- Oerbeck, B., Stein, M. B., Pripp, A. H. & Kristensen, H. (2015). Selective mutism: follow-up study 1 year after end of treatment. *European child & adolescent psychiatry*, 24(7), 757–766. <https://doi.org/10.1007/s00787-014-0620-1>
- Oerbeck, B., Stein, M. B., Wentzel-Larsen, T., Langsrud, Ø. & Kristensen, H. (2014). A randomized controlled trial of a home and school-based intervention for selective mutism – defocused communication and behavioural techniques. *Child and adolescent mental health*, 19(3), 192–198. <https://doi.org/10.1111/camh.12045>
- Petermann, F. (1996a). Einzelfallanalyse. München: Oldenbourg.
- Petermann, F. (1996b). *Einzelfalldiagnostik in der klinischen Praxis* (3., neu ausgestattete Aufl.). Weinheim: Beltz Psychologie-Verl.-Union.
- Pfaff, H., Neugebauer, E., Glaeske, G. & Schrappe, M. (Hg.). (2017). *Lehrbuch Versorgungsforschung: Systematik – Methodik – Anwendung* (2., vollständig überarbeitete Auflage). Stuttgart: Schattauer.
- Pionek Stone, B., Kratochwill, T. R., Sladeczek, I. & Serlin, R. C. (2002). Treatment of selective mutism: A best-evidence synthesis. *School Psychology Quarterly*, 17(2), 167–190. <https://doi.org/10.1521/scpq.17.2.168.20857>
- Reichert, M. & Genoud, P. A. (2015). *Einzelfallanalysen in der psychosozialen Forschung und Praxis*. Weitraamsdorf: ZKS-Verlag.
- Riley, G. D. (2009). *Stuttering Severity Instrument* (4th edition). Austin (TX): Pro-Ed.
- Sandrieser, P. & Schneider, P. J. (2015). *Stottern im Kindesalter* (4. Aufl.). Stuttgart: Thieme.
- Schrappe, M. [M.] & Pfaff, H. [H.] (2016). Versorgungsforschung vor neuen Herausforderungen: Konsequenzen für Definition und Konzept. *Gesundheitswesen (Bundesverband der Ärzte des Öffentlichen Gesundheitsdienstes)*, 78(11), 689–694. <https://doi.org/10.1055/s-0042-116230>
- Standart, S. & Le Couteur, A. (2003). The Quiet Child: A Literature Review of Selective Mutism. *Child and adolescent mental health*, 8(4), 154–160. <https://doi.org/10.1111/1475-3588.00065>
- Starke, A. & Subellok, K. (2016). *Dortmunder Mutismus Screening DortMuS-Schule: Manual*. <http://sk.tu-dortmund.de/dortmus/DortMuS-Schule.pdf>
- Starke, A. & Subellok, K. (2018). *Dortmunder Mutismus Screening DortMuS-Kita*. <http://www.sk.tu-dortmund.de/dortmus/DortMuS-Kita.pdf>
- Starke, A., Subellok, K. & Pickhinke, I. (2018). Förderung selektiv mutistischer Kinder im schulischen Kontext – Ergebnisse aus zwei kontrollierten Einzelfallstudien. In T. Jungmann, B. Gierschner, M. Meindl, S. Sallat (Hg.), *Sprachheilpädagogik aktuell: Band 3. Sprach- und Bildungshorizonte: Wahrnehmen – Beschreiben – Erweitern* (S. 355–361). Idstein: Schulz-Kirchner.
- Subellok, K., Katz-Bernstein, N., Bahrfeck-Wichitill, K. & Starke, A. (2012). DortMuT (Dortmunder Mutismus-Therapie): eine (sprach)therapeutische Konzeption für Kinder und Jugendliche mit selektivem Mutismus. *L.O.G.O.S. Interdisziplinär*, 20(2), 84–94.
- Viana, A. G., Beidel, D. C. & Rabian, B. (2009). *Selective mutism: a review and integration of the last 15 years*. *Clinical psychology review*, 29(1), 57–67. <https://doi.org/10.1016/j.cpr.2008.09.009>
- Wissenschaftsrat. (2012). *Empfehlungen zu hochschulischen Qualifikationen für das Gesundheitswesen*. Köln: Geschäftsstelle des Wissenschaftsrates.
- Yin, R. K. (2014). *Case study research: Design and methods* (5. edition). Los Angeles, London, New Delhi, Singapore, Washington (DC): SAGE.
- Zakszeski, B. N. (2016). Reinforce, shape, expose, and fade: a review of treatments for selective mutism (2005–2015). *School mental health*, 9, 1–15. <https://doi.org/10.1007/s12310-016-9198-8>

Zum Autor und den Autorinnen

Prof. Kohler ist Dozent an der Interkantonalen Hochschule für Heilpädagogik (HfH) in Zürich. Dort lehrt er schwerpunktmäßig zu Redeflussstörungen, Forschungsmethoden, Diagnostik, Entwicklungspsychologie und Lernpsychologie. Von 2016 bis 2019 untersuchte er den Einsatz des RedeflussKompass und StotterKompass in der sprachtherapeutischen Praxis.

Prof. Starke lehrt und forscht an der Universität Bremen im Bereich der inklusiven Pädagogik mit dem Schwerpunkt Sprache. Einer ihrer Forschungsschwerpunkte ist dabei der selektive Mutismus, wozu sie an der TU Dortmund Einzelfallstudien durchführte.

Prof. Kohmäscher lehrt an der FH Münster professionsübergreifend Therapiewissenschaften und Forschungsmethoden. Seit 2018 leitet sie die vom Gemeinsamen Bundesausschuss geförderte Versorgungsstudie PMS KIDS zur Wirksamkeit ambulanter Stottertherapien unter Alltagsbedingungen.

Korrespondenzadresse

Prof. Jürgen Kohler
Interkantonale Hochschule für Heilpädagogik (HfH)
E-Mail: juergen.kohler@hfh.ch