



Studienabschlussarbeiten

Fakultät für Psychologie und
Pädagogik

Feselmayer, Fabian:

Evidenzbasierte Entwicklung des Ganztags. Eine
Wissenschafts-Praxis-Kooperation mit einem externen
Kooperationspartner

Eine Mixed-Method Evaluation des Programms „Eine
Note besser!“

Masterarbeit, Sommersemester 2024

Gutachter*in: Eckert, Thomas

Fakultät für Psychologie und Pädagogik

Lehrstuhl für Allgemeine Pädagogik und Bildungsforschung

Pädagogik mit Schwerpunkt Bildungsforschung und Bildungsmanagement

Ludwig-Maximilians-Universität München

<https://doi.org/10.5282/ubm/epub.122072>

Inhaltsverzeichnis

Abbildungsverzeichnis:.....	3
Tabellenverzeichnis:	3
1. Einleitung.....	4
2. Theorie.....	6
2.1 Allgemeine Einführung.....	6
2.2 Ganzttag	6
2.3 Bildungsgerechtigkeit.....	11
2.4 Qualität im Ganzttag.....	12
2.5 Stiftungen.....	13
2.6 Lehrerüberzeugungen und -handeln:.....	14
2.7 Kooperation zwischen Ganzttag und externen Kooperationspartnern	16
2.8 Evaluation	17
3. Evaluation des Programms „Eine Note besser!“	20
3.1 Beschreibung des Programms	20
3.2 Chronologischer Verlauf der Evaluation	22
3.3 Evaluationsziele, Evaluationsmodell & Fragestellungen	23
3.4 Konzeption der Evaluation.....	25
4. Quantitative Analyse	27
4.1 Fragestellungen & Hypothesen	27
4.2 Stichprobe & Überblick Testungen	28
4.3 Erhebungsinstrumente.....	30
4.4 Methoden.....	33
4.5 Deskriptive Analysen	34
4.5.1 Soziodemographische und schul- und programmbezogene Daten	35
4.5.2 Leistungsrelevante Daten	35
4.5.3 Individuelle Lernverläufe.....	39
4.6 Inferenzstatistische Analysen.....	40
4.6.1 Diskrepanzvergleiche zwischen den ELFE-Untertests	40
4.6.2 Schülerleistungen über die Zeit	43

4.6.3 Veränderungen in der motivationalen Orientierung.....	46
5. Qualitative Analyse	47
5.1 Fragestellungen	47
5.2 Theoretische Grundlagen	48
5.2.1 Forschungsprogramm Subjektive Theorien	48
5.2.2 Auswertung Qualitative Inhaltsanalyse nach Mayring	51
5.3 Empirische Untersuchung	53
5.3.1 Erhebungsmethode und Stichprobe	53
5.3.2 Datenerhebung.....	54
5.3.3 Interviewleitfaden & Forschungsethik	54
5.3.4 Festlegen der Kategorien für den Leitfaden.....	56
5.3.5 Datenaufbereitung	58
5.4 Ergebnisse.....	58
5.4.1 Programm.....	59
5.4.2 Kontext.....	65
5.4.3 Personal.....	67
5.4.4 Evidenzbasierung	72
6. Rekonstruktion der Programmtheorie	80
7. Diskussion	82
7.1 Summative Evaluation	82
7.2 Formative Evaluation	87
7.3 Limitationen der Arbeit	89
7.4 Empfehlungen	90
8. Fazit.....	96
Literaturverzeichnis:.....	98
Anhang	104
Zu erhebende Konstrukte:.....	104
Schülerprofil:.....	105
Einverständniserklärung qualitative Interviews:.....	106
Leitfaden für das Fachpersonal:.....	108
Leitfaden für die Lehrkraft:	110
Erweiterter Leitfaden für die Programmleitung.....	111

Abbildungsverzeichnis:

Abbildung 1 Übersicht über die Funktionen des Ganztags 8

Abbildung 2 Wirkungslogik von ENB 22

Abbildung 3 Chronologischer Verlauf und Meilensteine des Evaluationsprozesses 23

Abbildung 4 Mittelwerte und Standardabweichung der Gesamtwerte der ELFE-Testergebnisse 36

Abbildung 5 Mittelwerte und Standardabweichung der SELLMO-Testergebnisse 36

Abbildung 6 Mittelwert und Standardabweichung der Noten der gesamten Stichprobe 37

Abbildung 7 Mittelwerte und Standardabweichungen der Noten der bereinigten Stichprobe 38

Abbildung 8 T-Werte der einzelnen Kinder über die vier Messzeitpunkte 40

Abbildung 9 Signifikante Diskrepanzvergleiche 42

Abbildung 10 Box-Plots zur Identifikation von Ausreißern der ELFE-Testergebnisse 43

Abbildung 11 Box-Plots mit signifikanten Post-hoc-Tests der ELFE-Testergebnisse 45

Abbildung 12 Rekonstruierte Programmtheorie ausgehend von den qualitativen Interviews 81

Tabellenverzeichnis:

Tabelle 1 Evaluationsmodell dieser Arbeit 24

Tabelle 2 Fragestellungen dieser Evaluation 25

Tabelle 3 Fragestellungen & Hypothesen der quantitativen Analyse 28

Tabelle 4 Schüleranzahl, Geschlecht und zuhause gesprochene Sprachen in den Stichproben 29

Tabelle 5 Erhebungszeitpunkte der quantitativen Längsschnittdaten 30

Tabelle 6 Einzelgeförderte Kinder 35

Tabelle 7 Deskriptive Statistiken der Übertrittsnoten auf weiterführende Schulen 39

Tabelle 8 Ergebnisse des Shapiro-Wilk-Tests der ELFE-Testergebnisse 44

Tabelle 9 Post-hoc-Analysen der ELFE-Testergebnisse 45

Tabelle 10 Ergebnisse der Shapiro-Wilk-Tests der ELFE-Testergebnisse - ohne Ausreißer 46

Tabelle 11 Post-hoc-Analysen der ELFE-Testergebnisse - ohne Ausreißer 46

Tabelle 12 Ergebnisse des Shapiro-Wilk-Tests der SELLMO-Testergebnisse 47

Tabelle 13 Ergebnisse der t-Test für die berechneten parametrischen Tests der SELLMO-Testergebnisse 47

Tabelle 14 Dauer der qualitativen Interviews in Minuten 54

1. Einleitung

Bildung besitzt stets auch eine ethische Dimension. Vor diesem Hintergrund lässt sich argumentieren, dass die Gesellschaft Bildung als einen gesellschaftlichen Auftrag betrachtet. Dies wird insbesondere damit begründet, gesellschaftliche Stabilität, Weiterentwicklung sowie Überlebensfähigkeit durch Eingliederung der jüngeren Generationen in die Gesellschaft zu gewährleisten, weshalb Bildung gesellschaftlich stark institutionalisiert wurde. Des Weiteren wird das Vermitteln von Bildung oder das Schaffen von Bildungsgelegenheiten gesellschaftlich als notwendig zur Selbstentfaltung moralisch begründet. Aus dieser Argumentation heraus lässt sich die Stellung und Wichtigkeit des demokratischen und gesellschaftlichen Diskurses darüber, was Bildung ist und wie sie institutionell gefördert werden soll, begründen.

Damit einhergehen allerdings immer auch ethische Grundfragen, die entweder konsensual explizit oder implizit angenommen werden oder aber auf deren Grundlage sich unterschiedliche Argumentationsstränge im gesellschaftlichen Diskurs begründen lassen. Dies lässt sich auch auf die konkreten pädagogischen Entscheidungen von Lehrenden übertragen, die u. a. den eigenen Bildungsauftrag, die Bildungsnotwendigkeit des Kindes sowie den gerechten Mitteleinsatz ethisch begründen müssen, wenn notwendig. Um nun politisch oder wissenschaftlich über jene ethischen Dimensionen nachzudenken, werden gegenstandsbezogene und als plausibel eingeschätzte moralische Urteile auf Wahrheit oder Falschheit hin überprüft (vgl. Fischer, 2020).

Zum Beispiel: „Das Kind zum Lernen zu zwingen ist moralisch erlaubt“, „Ich erwarte vom Kind, dass es fleißig lernt und dass es meine intendierten Versuche der Wissensvermittlung als Angebot nutzt, sich selbst zu entwickeln, ist moralisch gerechtfertigt“ oder aber „Dass Kinder diejenigen Inhalte erwerben müssen (z. B. die Inhalte des staatlich festgelegten Curriculums), die ich als Bildung definiere, ist moralisch legitim“. Anhand dieser (impliziten) moralischen Urteilen fächert sich ein ganzer politischer, wissenschaftlicher oder auch gesellschaftlicher Diskurs auf, der sich damit beschäftigt, wie auf Grundlage dieser Urteile möglichst eine „effektive“ und von den effektiven Lösungen die „effizienteste“ gefunden werden kann.

Fischer (2020) umschreibt ähnliche Situationen und stellt daraufhin die Frage, ob sich die Betroffenen – hier also die Kinder oder auch die Lehrenden – tatsächlich diese abstrakten moralischen Fragen stellen. Dabei wendet er ein, dass Moral immer nur etwas mit dem Allgemeinen zu tun hat und sich somit nicht auf einen einzelnen Menschen, sondern auf alle Menschen bezieht und teilweise illusorisch unterstellt wird, dass alle Menschen nach diesem definierten Allgemeinen handeln (Fischer, 2020). Als Antwort formuliert er den Ansatz, dieses Allgemeine durch das Individuelle zu ersetzen, denn das Allgemeine verfehlt immer die Perspektive des Individuellen (vgl. Fischer, 2020). In anderen Worten: Fragen sich Lernende tatsächlich, warum sie zur Schule gehen, warum sie fleißig lernen sollen oder warum sie (noch) nicht gebildet sind? Es lässt sich vorstellen, dass auf Nachfrage Dinge geäußert werden würden, wie: „Ich gehe zur Schule, weil ich muss, damit meine Eltern stolz auf mich sind, ich später einen Beruf ergreifen kann usw.“ Hierbei ließen sich somit diese allgemeinen Kategorien, die im gesellschaftlichen Diskurs formuliert werden und an das Kind gestellt werden, wiederfinden. Sollte dies der Fall sein, so internalisiert das Kind diesen äußeren moralischen Diskurs und richtet seine Sinnesfragen danach aus.

Im Gegensatz dazu argumentiert Fischer (2020) für die Einnahme einer individuellen Perspektive, bei der konkret betrachtet wird, mit welchen Motiven und mit welchen Erlebensformen das Individuum in einer konkreten Situation instantiiert ist und warum es sich so verhält, wie es sich verhält. Wenn man diese Argumentation auf den Bildungsdiskurs überträgt, ergäbe sich daraus ein völlig anderes ethisches Rahmenmodell von Bildung. Bildung würde hier von individueller Seite aus gedacht werden – und nicht von gesellschaftlicher. Dabei lässt sich von Grund auf die Frage stellen, inwiefern von der Gesellschaft festgelegte Bildungsinhalte, -ziele, -methoden und –zeiten sich mit der Idee einer vom Individuum gedachten Bildung vereinbaren lassen und inwiefern es tatsächlich die Gesellschaft sein soll, welche den Bildungsweg dem Individuum vorschreibt – mit begrenzter Auswahl aus ungleichwertigen Alternativen.

Dies widerspricht allerdings nicht dem Argument, dass die Gesellschaft Bildungsinstitutionen schaffen kann und/oder soll, sondern lediglich, inwiefern es die Gesellschaft ist, die ein Bildungsverständnis vorgibt und anschließend entsprechend institutionalisiert. In diesem Sinne lassen sich bereits Bildungsreformen identifizieren, die in diese Richtung agieren, beispielsweise die bisherigen Reformen bezüglich der Ganztagschule. Diese basieren auf verschiedenen Argumentationsweisen, nach denen die bisherigen Strukturen verändert oder neue Strukturen geschaffen werden. Durch die Integration von gesellschaftlichen, außerschulischen Organisationen in das System Schule und der gleichzeitig fehlenden curricularen Vorgaben wird ein Raum für die Gesellschaft geschaffen, partizipativer Teil des Bildungssystems zu sein und Bildung nach anderen Definitionsansätzen sowie Zugängen zu gestalten. Andererseits bedeutet dies auch, dass diese Gestaltungsmöglichkeiten u. a. von denjenigen Akteursgruppen genutzt werden können, die die entsprechenden finanziellen und Machtmittel besitzen. Dies soll insbesondere in Kapitel 2.5 kritisch diskutiert werden.

Im Zuge dieser Arbeit wird das Programm “Eine Note besser!” (ENB) vorgestellt. Hierbei handelt es sich um ein außerschulisches Programm, welches durch die Stiftung “Gesellschaft macht Schule” (GmS) und die Castringius Kinder & Jugend Stiftung konzipiert und umgesetzt wird. Ziel des Programms ist einerseits die individuelle Förderung des Kindes – ausgehend von dessen Interessen und Bedürfnissen - sowie andererseits gezielt jene Kinder zu fördern, die aus Familien stammen, die gesellschaftlich benachteiligt sind, wodurch sich gleichzeitig Benachteiligungen innerhalb des Bildungssystems ergeben. Die Stiftung tritt somit für jene Familien und deren Kinder ein, die das Bildungssystem vernachlässigt und die nicht im gesellschaftlichen Diskurs selbst präsent sind. Um diesem Ziel gerecht zu werden, geht ENB den Weg des Aufbaus eines engen und langfristigen Beziehungsverhältnisses zwischen Pädagogin oder Pädagoge zum Kind, des Eingehens auf die kindlichen Bedürfnisse und Sinnfragen sowie der Unterstützung bei der selbstgewählten Entfaltung und Auseinandersetzung mit der Welt.

Aufgabe dieser Arbeit ist es nun, einerseits zu einem summativen Urteil zu kommen, mit welcher Güte ENB seine selbstgesetzten Ziele erfüllt und andererseits, welche Programmverbesserungen sich empirisch und ausgehend von der wissenschaftlichen Fachliteratur skizzieren lassen, um die Effektivität und Effizienz der Programmzielerreichung zu steigern. Da im Zuge der Evaluation kein Evaluationsbericht erstellt wird, ist diese Arbeit Teil des formativen Evaluationsprozesses und darüber hinaus ein Medium der Wissenschafts-Praxis-Kommunikation. Gleichzeitig ist deren Leserschaft nicht ausschließlich im universitären Kontext zu verorten. Aufgrund dessen wird seitens des Autors einerseits inhaltlich versucht,

möglichst dem Forschungsprozess zugrunde liegende Annahmen und Schlussfolgerungen zu explizieren und andererseits einen möglichst klaren und verständlichen Schreibstil anzuwenden, der sowohl den wissenschaftlichen als auch kommunikativen Kriterien dieser Arbeit gerecht wird. Allerdings lässt es sich nicht ganz vermeiden, teils komplexe Sätze zu formulieren und inhaltliches Vorwissen vorauszusetzen. Diese Arbeit versucht somit, einen Mittelweg zu gehen und sowohl den wissenschaftlichen als auch kommunikativen Kriterien zu erfüllen.

2. Theorie

2.1 Allgemeine Einführung

Da die Fragestellungen durch Aushandlungsprozesse im Verlauf der Evaluation formuliert wurden und nicht als Schlussfolgerungen ausgehend von Theorien, ergeben sich diese somit nicht schlüssig aus der Theorie. Somit soll der nachfolgende Theorieteil Evaluation an sich und ihren Evaluationsgegenstand theoretisch fundieren sowie wichtige Begrifflichkeiten, welche die Rahmenbedingungen der Evaluation beschreiben, explizieren.

Bei dieser Evaluation handelt es sich um eine Mixed-Method-Evaluation, welche sich sowohl quantitativer als auch qualitativer Forschungsmethoden bedient, um den Evaluationsgegenstand zu analysieren. Die theoretischen Konstrukte, die im Rahmen der quantitativen Analyse herangezogen werden, werden nicht im Theorieteil dargestellt, sondern zusammen mit den jeweiligen Erhebungsinstrumenten in Kapitel 4.3, da diese die theoretische Fundierung vorgeben und nicht seitens des Autors dieser Studie ausgewählt wurden. Diejenigen theoretischen Konstrukte, die im Rahmen der qualitativen Datenerhebung und -analyse verwendet wurden, um Leitfadencategorien und deduktive Kategorien festzulegen, werden lediglich zum Teil im Theorieteil angedeutet.

2.2 Ganzttag

Diese Evaluation findet im Kontext der Ganzttagsschule statt. Bei dem Evaluationsgegenstand handelt es sich um ein Programm eines für die Ganzttagsschule externen Kooperationspartners, dessen Rahmenbedingungen durch die Ganzttagsschule gegeben werden. Aufgrund dieser Einbettung in die Ganzttagsschule soll im Folgenden zunächst der Begriff der Ganzttagsschule und die damit zusammenhängende Qualitätsdebatte erläutert werden, da es das Ziel dieser Evaluation ist, (eine bestimmte) Qualität nachzuweisen oder zu indizieren.

Als Initiator des heutigen Diskurses über die Qualität des deutschen Bildungswesens gilt gemeinhin der "PISA-Schock" (Vogel, 2019), wobei insbesondere die Einführung und der Ausbau der Ganzttagsschule bereits seit den 70er-Jahren diskutiert wurde. Der Diskurs der letzten zwei Jahrzehnte über die Qualität des deutschen Bildungswesens wird allerdings im Gegensatz zu damals durch das Thema der Ganzttagsschule geprägt (KMK, 2002). Zentral in diesem Diskurs ist die Annahme, dass die Einführung und der Ausbau der Ganzttagsschule das in der Schule situierte Lernen verändern und somit auch verbessern kann (Schaller, 2020). Dieser Diskurs findet sowohl auf wissenschaftlicher als auch praktischer Ebene statt.

Definition von Ganztag:

Zur definitorischen Bestimmung des Begriffs Ganztagschule wird sich in dieser Arbeit an der Definition der Kultusministerkonferenz (KMK, 2018) aus 2004 orientiert: Ganztagschulen sind Schulen, die an mindestens drei Schultagen ein Unterrichtsangebot von sieben Stunden sowie an jedem Schultag den teilnehmenden Schülerinnen und Schülern ein Mittagessen offerieren. Die Angebote werden unter Aufsicht der Schulleitung konzipiert und mit dem bestehenden Unterrichtsangebot verbunden. In 2015 wurde die Rolle der Schulleitung relativiert und es wurde beschrieben, dass die Schulleitung lediglich mit einem außerschulischen Träger auf Basis eines gemeinsamen pädagogischen Konzepts kooperieren müsse und eine Mitverantwortung für das Angebot besitze, um als offene Ganztagsangebote zu zählen (KMK, 2023). Der ganztagsschulische Unterricht wird in diesem Sinne durch die KMK (2015) in drei Bereiche geteilt: Den regulären Pflichtunterricht, spezifische pflichtunterrichtliche Ergänzungen sowie außerunterrichtliche Angebote. Bei ENB handelt es sich sowohl um außerunterrichtliche Angebote als auch pflichtunterrichtliche Ergänzungen.

An dieser Definition wird erkennbar, dass Ganztagschule lediglich anhand des äußeren Kriteriums des Angebots, weder inhaltlich noch methodisch definiert wird – einzig mit der Voraussetzung, dass das Angebot in irgendeiner Form mit dem bestehenden Unterrichtsangebot verknüpft sein soll. Als Konsequenz dieser inhaltlichen Offenheit existiert eine große Angebotsvielfalt zwischen den Ganztagschulen, die unterschiedliche Foki besitzen sowie auf verschiedene Weise in das schulische Angebot integriert sind (Schulz-Gade, 2015). Hieraus ergibt sich wiederum die Schwierigkeit, einheitliche Qualitätsstandards zu formulieren, die für alle Angebote im Ganztag zu gleichen Teilen gelten. Dies lässt sich bereits auf einen großen Teil des Bildungssystems übertragen; im Jahr 2023 besaßen ca. 70 % aller Schulen in Deutschland einen Ganztagsbetrieb (KMK, 2023).

An die Ganztagschule werden aus Gesellschaft und Politik verschiedene Ansprüche gestellt, die insbesondere gesellschaftliche Schlüsselprobleme aufgreifen. So wird u. a. von der Ganztagschule erwartet, quantitativ und qualitativ bessere individuelle Förderung der Kinder zu gewährleisten, sozial bedingte Bildungsungleichheiten zu nivellieren sowie für Eltern eine bessere Vereinbarkeit von beruflichen und familiären Verpflichtungen (Züchner & Fischer, 2011). Daraus lässt sich ableiten, dass nicht nur reine ökonomische Leistungsmotive hinter den Ansprüchen an die Ganztagschule stehen, sondern weitere Dimensionen wie das eines Bildungsideals oder auch einer gerechten Gesellschaft. Nichtsdestotrotz stellen Effizienz- und Effektivitätskalküle einen großen Bereich der aktuellen Debatte dar (Ditton, 2010). In diesem Kontext wird häufig die gesteigerte Lernzeit und eine größere (Unterrichts-)Angebotsvielfalt angegeben, um den höheren Nutzen der Ganztagschule zu begründen (Strietholt et al., 2015). Das bedeutet, es werden sowohl quantitativ als auch qualitativ orientierte Argumente geäußert, wobei hierbei nicht nur die Qualität gemeint ist, sondern auch die inhaltliche und methodische Ausrichtung des Angebots. Dies hängt damit zusammen, da durch die längere Schulzeit mehr Unterrichtsstunden in den Ganztag integriert werden können und somit einerseits das bereits bestehende Angebot aufrechterhalten und andererseits in den zusätzlichen Unterrichtsstunden ein weiteres, qualitativ verschiedenes Angebot konzipiert und realisiert werden kann.

Inhaltlich werden von der Schulpolitik an die Ganztagschule ebenfalls allgemeine Anforderungen gestellt (Jugendinstitut, 2016): Es sollen insbesondere kognitive und soziale

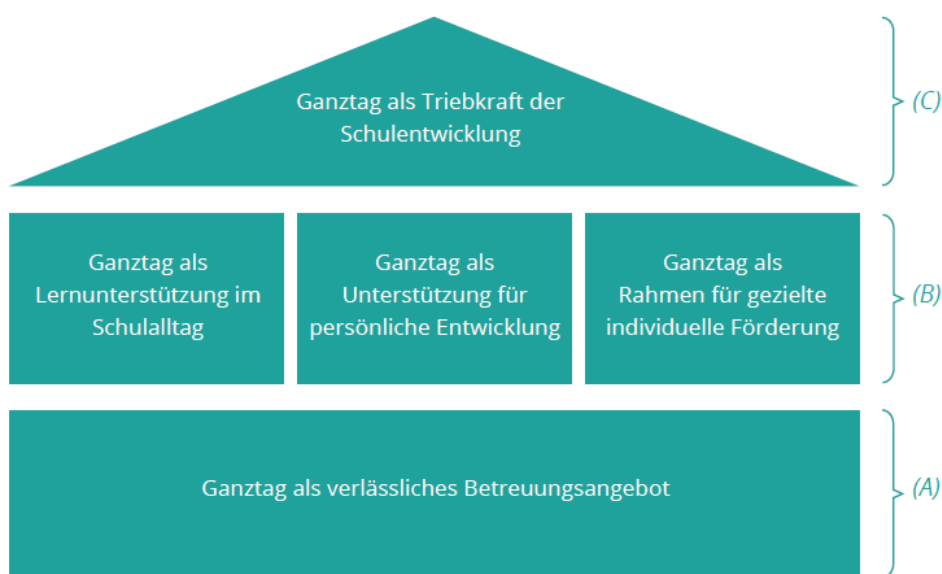
Kompetenzen durch eine nachhaltige und ganzheitliche Bildung entwickelt werden. Da für Ganztagsangebote keine inhaltlichen Vorgaben gemacht werden, sondern lediglich seitens der Gesellschaft und Politik Ansprüche formuliert werden, stellt sich die Frage, inwiefern insbesondere außerunterrichtliche Angebote diese Lücken im Lehrplan zu füllen vermögen. In diesem Sinne lässt sich argumentieren, dass Kritikerinnen und Kritiker des bestehenden Lehrplans und der dort vorkommenden Inhalte aufgrund dieser inhaltlichen Offenheit der Ganztagsangebote ein neues Vehikel ihrer Kritik bekommt und somit verschiedene Kritikerinnen und Kritiker bedient werden können, die ihre Erwartungen und Ansprüche in die Angebote projizieren können. Wird dieser Argumentation gefolgt, so kann die Öffnung des Ganztags für außerschulische Akteurinnen und Akteure auch in diesem Sinne interpretiert werden.

Funktionen des Ganztags:

In StEG (Studie zur Entwicklung von Ganztagsschulen) werden die Funktionen des Ganztags folgendermaßen beschrieben, die in Abbildung 1 dargestellt und aufgezählt werden (Jugendinstitut, 2019):

- A. Die dem Ganztag attribuierte Betreuungsfunktion stellt das Fundament für alle weiteren, darauf aufbauenden Funktionen dar.
- B. Des Weiteren kommt dem Ganztag die Funktion der Unterstützung und individuellen Förderung des Lern- und Entwicklungsprozesses der Schülerinnen und Schüler zu.
- C. Die koordinierte Konzeption und Umsetzung der Unterstützungs- und individuellen Förderungsangebote führt zur Etablierung von Strukturen und (multiprofessioneller) Kooperation zwischen den schulischen und außerschulischen Akteurinnen und Akteuren.

Abbildung 1 Übersicht über die Funktionen des Ganztags



Jugendinstitut, 2019, S. 6

Auf gesamter Schulebene zeichnen sich allerdings über die letzten 20 Jahre im Schnitt anhand der PISA-Ergebnisse (z. B. Lewalter et al., 2023) oder aber des IQB-Bildungstrend (Stanat et al., 2022) keine signifikanten Verbesserungen, sondern eher Verschlechterungen, insbesondere für sozial benachteiligte Kinder, ab. Die Ergebnisse dieser Large-Scale-Assessments werden größtenteils herangezogen, um auf mangelnde Qualität in den Schulen und darunter auch Ganztagschulen hinzuweisen und die gestellten politischen Ansprüche als nicht erfüllt zu betrachten. So wurden in der dritten Startkohorte von NEPS (National Educational Panel Study) die Kompetenzverläufe der Schülerinnen und Schüler im Lesen und Mathematik u. a. im Vergleich zwischen Ganz- und Halbtagsschule untersucht sowie zusätzlich die Angebotsnutzung und -bewertung der Ganztagskinder erfragt (Linberg et al., 2018): Hierbei konnten keine Effekte weder der Ganztagschule noch der Angebotsnutzung auf die Kompetenzentwicklung der Schülerinnen und Schüler gefunden werden (Linberg et al., 2018). Dies widerspricht insbesondere der Annahme, dass durch die Einführung und den Ausbau der Ganztagschulen die Schulqualität gesteigert wurde bzw. werden kann.

Allerdings gibt es auch Studien, die der Teilnahme an außerunterrichtlichen Ganztagsangeboten konkrete Wirkungen zuschreiben: So entdeckte u. a. StEG, dass die Teilnahme am Ganzttag das Sozialverhalten, die Schulfreude sowie die Noten der Schülerinnen und Schüler verbessern und gleichzeitig die Gefahr der Klassenwiederholung reduzieren kann (Jugendinstitut, 2016). Aber: Dies gilt nur, wenn die Qualität der Angebote gut ist, diese ausreichend genutzt werden und das Beziehungsklima der Schule stimmt (Jugendinstitut, 2016). Gleichzeitig zeigten Linberg et al. (2018) anhand der NEPS-Daten auf, dass lediglich dann ein Effekt der außerunterrichtlichen Förderung auf die kindliche Kompetenzentwicklung gefunden werden konnte, wenn das Ganztagsangebot durch die Kinder positiv bewertet wurde. Des Weiteren untersuchte StEG, inwieweit die Angebote und der Unterricht explizit miteinander verknüpft seien und kam hierbei zu dem Urteil, dass dies in 50 % aller Ganztagschulen nicht der Fall sei (Jugendinstitut, 2016). Aus dieser Nicht-Erfüllung der an die Ganztagschule gestellten Ansprüche ergibt sich nun ein (wissenschaftlicher) Diskurs der Ursachenanalyse.

Gleichzeitig lässt sich ein Phasenwechsel hinsichtlich der Ausbaubemühungen des Ganztages konstatieren, insofern als von dem vornehmlich quantitativ orientierten Ausbau von Ganztagschulen nun die qualitativ orientierte Weiterentwicklung der Angebotsqualität der bestehenden Ganztagschulen in den Fokus gerät (Rauschenbach, 2012). In diesen Kontext lässt sich auch diese Evaluation einordnen. Problematisch an dieser (wissenschaftsbasierten) Ursachenanalyse ist, dass eine Vielzahl an Faktoren empirisch herangezogen werden kann und wird, um die mangelnde Qualität in den Ganztagschulen bzw. die erwartungswidrigen Leistungsergebnisse zu begründen (Tillmann et al., 2018). Hierbei treten teilweise auch Widersprüchlichkeiten zwischen den gefundenen Ergebnissen auf (z. B. Wiere, 2011). Ein hierbei geäußelter Argumentationsstrang bezieht sich auf das Fehlen klarer Vorgaben sowie Qualitätskriterien seitens der Bildungspolitik (Rauschenbach, 2012), worin sich die Angebotsvielfalt und darauf aufbauend fehlende Steuerungsmechanismen begründen. Dies kann einer der Gründe sein, warum einerseits auf bildungspolitischer Ebene kein Qualitätsstandard etabliert und garantiert werden kann und andererseits wieso möglicherweise die Angebote nicht mit den zur Beurteilung des Bildungssystems herangezogenen Qualitätsindikatoren konvergieren. Gängler (2011) bezieht sich wiederum auf das in den Angeboten tätige Personal, welches neben Lehrkräften auch weiteres (pädagogische) Fachpersonal mit unterschiedlichen disziplinären Hintergründen umfasst. Hierdurch entstehen in der Schule neue Formen der (multiprofessionellen) Kooperation

zwischen den schulischen Akteurinnen und Akteuren. Gleichzeitig lässt sich vorstellen, dass dieses neu hinzukommende Fachpersonal eigene Bildungsverständnisse, methodisch-didaktische Vorgehensweisen und Rollenverständnisse hat, die zum Teil biographisch und zum anderen Teil situativ bestimmt werden.

In StEG wird die ausbleibende Wirkung von Ganztagsangeboten folgendermaßen begründet (Jugendinstitut, 2019): Zunächst wird kritisch betrachtet, dass die Intensität der Angebote unzureichend sein könnte, da diese lediglich ein- bis zweimal die Woche eine geringe Anzahl an Schülerinnen und Schülern betreffen. Des Weiteren wird in Erwägung gezogen, dass bestimmte Angebote nur durch diejenigen Schülerinnen und Schüler genutzt werden, die schulisch bereits gut dastehen und deswegen ihre Zeit und Energie in die Ganztagsangebote investieren können. Darüber hinaus nennt StEG ebenfalls die mangelnde Qualität der Ganztagsangebote als bedeutende Ursache für ausbleibende Wirkungen und schlägt gleichermaßen vor, Rahmenvorgaben sowie Curricula zu entwickeln, an welche sich die Ganztagsangebote orientieren müssen (Jugendinstitut, 2019). StEG zieht dabei das Fazit, dass die Fördermaßnahmen im Ganztage zeitlich intensiver, systematischer, stärker basierend auf Forschungsbefunden und möglichst standardisiert konzipiert und realisiert werden sollten (Jugendinstitut, 2019): Insbesondere empfehlen die Autorinnen und Autoren eine intensive Zusammenarbeit zwischen den ganztage schulischen Akteurinnen und Akteuren und denen der Wissenschaft, insbesondere bei der Konzeption der Angebote. Denn individuelle Angebote zur Förderung von fachlichen und überfachlichen Kompetenzen bedürfen einer Systematizität und ausgearbeiteten Konzeption, um qualitativ zu sein (Jugendinstitut, 2019). StEG kritisiert dabei, dass solche systematische Entwicklungsarbeit noch zu selten stattfindet (Jugendinstitut, 2019): Gleichzeitig kann die Wissenschaft den Rahmen der Konzeption und der Realisierung der durch sie begleiteten Ganztage angeboten nutzen, um partizipativ den Prozess zu begleiten und Erkenntnisse zu gewinnen.

Fazit:

Somit lässt sich zusammenfassen, dass der Fokus im Kontext der aktuellen wissenschaftlichen Betrachtung der Ganztage schule stärker auf Qualitätsfaktoren und deren Wirk- und Interaktionsmechanismen in allgemeinen und konkreten Kontexten und Situationen gerichtet wird. Diesen Fokus will diese Arbeit ebenfalls übernehmen, indem als Teil der formativen Evaluation explizit qualitätsrelevante Konstrukte und die Einbettung des Programms in dessen (Schul-)Kontext beleuchtet werden (vgl. Mandl & Hense, 2007). Aufgrund der beschriebenen heterogenen Ausprägungsformen des Ganztage und dessen Angeboten sowie den ebenso heterogenen zugrundeliegenden Qualitätsmodellen und Wirkfaktoren wird der Vergleich zwischen den verschiedenen Angeboten und deren Outputs und Outcomes erheblich erschwert. Dies erschwert insbesondere aufgrund der mangelnden Datengrundlage aufseiten der Wissenschaft sowie aufseiten des Angebots den gemessenen Output und Outcome im Rahmen der Evaluation auf das Angebot zurückzuführen. Folglich sind Vergleiche zwischen verschiedenen Angeboten bzw. Programmen mit Vorsicht zu ziehen, wobei fraglich ist, welche Erkenntnis gewonnen und welche Schlüsse gezogen werden können, wenn zentrale Einflussfaktoren und Gegebenheiten nicht gegeben sind. Somit ist eines der Ziele dieser Studie – neben den konkreten Evaluationszielen – einen Beitrag zur wissenschaftlichen Datenlage über den Ganztage und dessen Angeboten zu liefern.

2.3 Bildungsgerechtigkeit

Wie bereits beschrieben, steht im wissenschaftlichen und politischen Diskurs über die Ganztagschule deren Qualität im Vordergrund. Da Qualität ein normativer Begriff ist, der keinen spezifischen Inhalt transportiert, muss der Begriff somit inhaltlich bestimmt werden. Das theoretische Konstrukt der „Bildungsgerechtigkeit“ kann und wird in diesem Kontext herangezogen, um die Qualität des Bildungssystems zu bewerten bzw. zu entwickeln. Da ENB sein Programm zu großen Teilen auf Basis der Chancen- und Bildungsgerechtigkeit begründet und fundiert, soll nun im Folgenden geklärt werden, was unter den Begriffen Bildungsgerechtigkeit sowie Chancengerechtigkeit verstanden wird: Rauschenbach (2012) beschreibt, dass eine chancengerechte Bildung eines der Hauptmotive war, welches historisch im Zusammenhang mit dem Ausbau des Ganztags geäußert wurde. Diese bezog sich auf den Ausgleich der ungleichen Chancenverteilung zwischen jenen Kindern, die aus „bildungsnahen“ und jenen, die aus „bildungsfernen“ Familien stammten (Rauschenbach, 2012). So lässt sich schlussfolgern, dass sich insbesondere die Nivellierung von sozialer (Bildungs-)Ungleichheit nach gerechtigkeitstheoretischen Argumenten häufig in Form von Chancen- oder Bildungsgerechtigkeit begrifflich im Diskurs um die Ganztagschule manifestierte. Als zentral fungiert hierbei die gerechtigkeitstheoretische Dimension der Nicht-Rechtfertigbarkeit von sozialer Ungleichheit aufgrund von leistungsfremden Merkmalen in modernen Gesellschaften, auf welcher sich die Debatte um Bildungsgerechtigkeit im Zusammenhang mit dem Ganztags tag situieren lässt (vgl. Ditton, 2013).

In der aktuellen Debatte wird der Begriff der Bildungsgerechtigkeit zuvorderst im Kontext der Verbesserung des bestehenden Bildungssystems herangezogen, insofern als damit politische Handlungsbedarfe identifiziert und begründet werden sollen (vgl. Brake & Büchner, 2012; Ditton, 2013). Dadurch werden Ausbaubemühungen damit gerechtfertigt, einen Beitrag zu jener Bildungsgerechtigkeit zu leisten. Dies lässt sich auch auf ENB übertragen, welche ebenfalls den Begriff der Chancengleichheit synonym heranziehen, um die Notwendigkeit und die Zielausrichtung ihres Programms zu begründen. Allerdings wird der Begriff der Bildungsgerechtigkeit oder auch Chancengleichheit in der wissenschaftlichen Literatur kritisch gesehen, wobei insbesondere die oftmals willkürliche Verwendung des Begriffs kritisiert wird (vgl. Ditton, 2013). Hierbei wird der Begriff der *Chance* von Ditton (2013) als immanenter Bestandteil des Begriffs der Bildungsgerechtigkeit betrachtet und somit scheint eine Abgrenzung zwischen Chancengleichheit oder Bildungsgerechtigkeit zunächst begrifflicher, aber nicht semantischer Natur. Ditton (2013) definiert allerdings nicht konkret, was unter dem Begriff Chancen zu verstehen ist. Er könnte sich darauf beziehen, Zugang zu Bildungsangeboten zu erhalten, die das Potenzial haben, eigene Defizite, die in der eigenen sozialen Herkunft begründet liegen, abzubauen. Die Zieldimension der Chancengleichheit sind nämlich nach Ditton (2013) die gesellschaftlichen Teilhabechancen. Das bedeutet, dass der Begriff der Chancengleichheit in Bezug auf das Bildungssystem konkrete Möglichkeiten zur Bildung beschreibt, die wiederum konkrete Möglichkeiten zur Teilhabe am gesellschaftlichen Leben führen. Folglich lässt sich somit feststellen, dass laut dieser Argumentation durch Ganztagschulen jene Chancen auf Bildung geschaffen oder vergrößert werden sollen, die wiederum zu einem Abbau herkunftsbedingter und bildungsrelevanter Ungleichheiten führen. Aufgrund der Unklarheiten bzgl. der Begriffsdefinition von Chancen erscheint es wichtig, im Kontext dieser Evaluation, das Begriffsverständnis von ENB zu identifizieren und daraus konkrete Zieldimensionen abzuleiten, die wiederum durch die Evaluation überprüft werden können.

2.4 Qualität im Ganzttag

Im Folgenden soll nun dargestellt werden, welche Arten von Qualitätsdefinitionen im Zuge dieser Evaluation herangezogen werden können, um die Programm- und Angebotsqualität zu bewerten. Dazu wird sich größtenteils auf StEG bezogen: In StEG wurden einerseits Schülerinnen und Schüler des Primar- und Sekundarbereichs befragt, wie sie die Qualität der Angebote wahrnehmen (Jugendinstitut, 2016): Hierbei bewerteten die Schülerinnen und Schüler die Qualität im Allgemeinen als sehr positiv und nannten in diesem Zusammenhang besonders deren Motivierungsqualität sowie die emotionale Anerkennung, die ihnen durch die Angebote zuteilwurde. Kritisch wurde allerdings die Qualität hinsichtlich der Unterstützung der Schülerinnen und Schüler zu (mehr) Autonomie gesehen (Jugendinstitut, 2016). Andererseits untersuchte StEG konkrete Einflussfaktoren, die auf dieses positive Schülerurteil hinsichtlich der Qualität wirken könnten (Jugendinstitut, 2016): Hierbei wurde u. a. von den Schülerinnen und den Schülern die Abwechslung vom Schulunterricht genannt, insofern das Angebot anders als der Schulunterricht gestaltet wurde. Gleichfalls wurde die bewertete Qualität in Zusammenhang mit den Partizipationsmöglichkeiten innerhalb des Programms sowie die durch das Programm zuteilwerdende individuelle Anerkennung genannt (Jugendinstitut, 2016). Die Autorinnen und Autoren von StEG urteilten demnach, dass die Schülerinnen und Schüler insbesondere dann die Angebotsqualität als gut bewerteten, wenn eine gute Passung zwischen deren Bedürfnissen/Interessen und dem Angebot besteht (Jugendinstitut, 2016). Falls die Angebotsqualität als positiv bewertet wurde, konnten positive Einflüsse auf die Motivation, das Selbstkonzept und das soziale Lernen der Schülerinnen und Schüler gefunden werden (Jugendinstitut, 2016). Gleichzeitig lassen sich auch Effekte auf die Testleistungen der Schülerinnen und Schüler finden, wenn eine Passung mit den Schülerinteressen vorliegt (Jugendinstitut, 2016). Allerdings konnten keine statistisch signifikanten Veränderungen der Lese- und naturwissenschaftlichen Kompetenzen im Vergleich zur Kontrollgruppe vorgefunden werden, auch nicht bei hoher Angebotsqualität oder intensiver Teilnahme (Jugendinstitut, 2016). Daraus lässt sich schließen, dass StEG keine konkreten Qualitätsmerkmale identifizieren konnte, um (statistisch signifikante) fachliche Kompetenzzuwächse – welche sich zum Teil in den Noten widerspiegeln sollten – innerhalb der Erhebungsspanne aufzeigen zu können, neben der Passung mit den Schülerinteressen.

Anknüpfend an die Angebotsqualität lässt sich die Frage zur Unterrichtsqualität definieren, da die konkreten Angebote auch als unterrichtliche Situationen beschrieben werden können. Diese Arbeit soll sich nach Ditton (2009) richten, welcher Unterrichtsqualität zweidimensional beschreibt: Einerseits beschreibt er Unterrichtsqualität als Übereinstimmung der Unterrichtspraxis mit bestimmten normativen Modellen. Andererseits beschreibt er als Unterrichtsqualität die empirisch feststellbaren Wirkungen und Effekte von Unterricht auf die Schülereigenschaften (Ditton, 2009). Hierbei soll sich allerdings an dem Begriff der Unterrichtsqualität als normative Vorgabe orientiert werden. Dies hat insbesondere den Grund, da die Literatur zu empirisch beobachtbaren Wirkungen und Effekte von Unterricht und die damit verbundenen Kontextbedingungen noch spärlich ist und somit noch keine belastbaren Daten herangezogen werden können, um Qualitätsdimensionen für Ganztagsangebote zu beschreiben (vgl. Jugendinstitut, 2019).

Für die Programmqualität sollen einerseits die Empfehlungen der Kultusministerkonferenz aus 2023 für ganztägige Bildungs- und Betreuungsangebote herangezogen werden, um einen normativen Bewertungsrahmen zu etablieren (vgl. KMK, 2023): Das Programm ermöglicht das Schaffen von motivierenden, inklusiven und bedürfnisgerechten Lern- und Freizeitgelegenheiten innerhalb der außerschulischen

Angebote sowie die Kooperation zwischen den außerschulischen und schulischen Akteursgruppen. Gleichzeitig lässt sich die Programmqualität als Übereinstimmung des Bildungsverständnisses zwischen außerschulischen und schulischen Akteursgruppen beschreiben (vgl. KMK, 2023). Andererseits beschreiben Floercke und Holtappels (2004) die Struktur- und Prozessqualität des Programms und dessen Angebote als entscheidend (vgl. Arnoldt, 2011): Diese wird beeinflusst durch drei Bereiche: Erstens das Programm mit seinen rechtlichen und fachlichen Zielvorgaben sowie der Kooperationsvereinbarung mit der jeweiligen Schule; zweitens die Implementationsstrukturen des Programms wie die verbindlichen Kooperationsprozesse sowie die verfügbaren Ressourcen und drittens das konkrete Kooperationsverhalten der Akteurinnen und Akteure vor dem Hintergrund deren Orientierungen und Interessen. Die Zieldimension Strukturqualität wird hier beschrieben als Qualifikation des Personals, institutionalisierte Kooperationsstrukturen, Schulstrukturfaktoren, Ressourcenausstattung sowie Fort- und Weiterbildungsmöglichkeiten. Die Prozessqualität wiederum wird als Gestaltungen der Handlungen wie Kooperationsstile, Koordination der institutionellen Einzelleistungen und Abstimmung des professionellen Handelns, Entscheidungs- und Partizipationsstrukturen sowie konzeptionelle Arbeit beschrieben. Dabei wird insbesondere – wie schon von KMK (2023) betont – die Verknüpfung von Angebot und schulischem Unterricht betont, da die außerschulischen Angebote meist lediglich additiv zum Ganztags konzipiert werden und nicht interaktiv (vgl. Arnoldt, 2011).

2.5 Stiftungen

Stiftungen sind nun diejenigen Akteurinnen, welche die Angebots- und Programmqualität definieren und realisieren sowie konkrete Rahmenbedingung festlegen – wie Personal, Zeit, Ressourcen, Organisationsform, Programmziele – welche Einfluss auf die Angebots- und Programmqualität haben können (vgl. Schulz-Gade, 2015). Neben der Frage, welche stiftungsbedingten Rahmenbedingungen allgemein für Ganztagsangebote vorliegen, stellt sich gleichermaßen die Frage, was Stiftungen überhaupt sind, welche Rolle sie im Bildungswesen haben und welche möglichen Ziele sie als außerschulische, nicht-staatliche Akteurinnen verfolgen. Diesen Fragen soll im Folgenden nachgegangen werden:

Gemeinnützige oder philanthropische Stiftungen – unter welche auch GmS sowie ein Teil seiner Finanzgeber fallen – werden als Teil des „nicht-profitorientierten“ oder auch „Dritten“ Sektors beschrieben (vgl. Koranyi, 2021). Diese Bezeichnung entstand ideengeschichtlich aus dem Diskurs, einen weiteren gesellschaftlichen Sektor zu öffnen, welcher weder profitorientiert noch steuerfinanziert ist (Etzioni, 1973). Koranyi (2021) definiert dem Dritten Sektor zugehörige Stiftungen folgendermaßen: Dem Dritten Sektor gehören jene organisierte, private (nicht-staatliche), selbstverwaltete und gemeinnützige Organisationen an, welche generierte Gewinne nicht an Shareholder ausschütten (vgl. Salamon & Anheier, 1997; Evers & Laville, 2004). Thümler et al. (2014) argumentieren ausgehend von dieser Beschreibung, dass der Einbezug dieses Dritten Sektors in das Bildungssystem eine der bedeutendsten (internationalen) Reformen der letzten Jahrzehnte sei. Im Zuge dieser Reformen wurde Schulen mehr Verantwortung übertragen, Bildung dezentraler zu gestalten, um einen Bildungsmarkt zu eröffnen, in welchem Schulen wettbewerbsmäßig ihr eigenes Profil bilden, um sich gegenüber anderen Schulen abzugrenzen (vgl. Yemini & Sagie, 2015): Dies bringt insbesondere die Möglichkeit der Öffnung der Schule für außerschulische Akteurinnen und Akteure mit sich, die zusätzliche (nicht-steuerfinanzierte und nicht-staatlich reglementierte) Ressourcen und Expertise mit in den Bildungsbereich einbringen und die schulische

Lebenswelt mit der außerschulischen Lebenswelt stärker miteinander verknüpfen (Eyal & Yarm, 2018; Meyer & Boyd, 2001). Dies begründet sich an der Erwartungshaltung, dass die Schule gemeinsam mit ihren Partnerinnen und Partnern gesellschaftliche Probleme aus verschiedenen Perspektiven angehen und somit einen Beitrag zum Allgemeinwohl leisten kann (Koranyi, 2021; Crosby & Bryson, 2010). Eine weitere Argumentation für den Einbezug von außerschulischen Akteurinnen und Akteuren bzw. konkret von Stiftungen ist, dass Schulen somit flexibler auf die spezifischen (lokalen) Kontextbedingungen reagieren können und gleichzeitig unter Einbezug dieser Kontextbedingungen neue Inhalte und Methoden Anwendung finden können, die zunächst nicht traditionell oder begrifflich mit Schule verbunden werden (Yemini et al., 2018; Kolleck, 2017).

Neben diesen Vorteilen werden allerdings ebenfalls die damit einhergehenden, teilweise unkontrollierten Einflussmöglichkeiten von Stiftungen kritisch betrachtet (Horvath & Powell, 2016; Brecher & Wise, 2008). So kritisieren Gautreaux und Delgado (2016), dass Stiftungen mit eigenen (verpflichtenden) Ausbildungsangeboten – die das von der Stiftung angestellte Fachpersonal auf ihre Tätigkeit vorbereiten soll – staatlich nicht-reglementierte Ausbildungsstrukturen parallel zu der (staatlich reglementierten) Fachkräfteausbildung schaffen. Darüber hinaus beschreiben Ball und Junemann (2012), dass Stiftungen Netzwerke aufbauen und verwenden, um ihre Interessen mit anderen Stiftungen zu bündeln und politisch sowie schulpolitisch Einfluss zu nehmen. Insbesondere lässt sich kritisch betrachten, dass die stiftungseigenen Ideologien somit ein gesellschaftlich legitimes Vehikel bekommen, um Bildungsinhalte zu bestimmen (Kolleck, 2017; Harrow, 2011). Koranyi und Kolleck (2018) definierten ausgehend von der bestehenden Literatur vier zentrale Rollenverständnisse von Stiftungen (vgl. Koranyi, 2021):

1. **Finanzielle Förderung und operative Tätigkeit:** Stiftungen als Kapitalgeber, die zusätzliche, nicht-steuerfinanzierte Mittel hinzugeben sowie selbst operativ tätig sind und somit den staatlichen Personalbedarf senken bzw. nicht erhöhen.
2. **Innovation:** Aufgrund der inhaltlichen Unabhängigkeit von demokratischen Legitimationsprozessen können Stiftungen Innovationen im Bildungssystem umsetzen, die ansonsten gescheitert wären.
3. **Interessenvertretung:** Stiftungen vertreten direkt oder indirekt bestimmte Interessen, wozu sie verschiedene Mittel einsetzen.
4. **Beziehungsaufbau und Netzwerkerin:** Durch den Aufbau von Beziehungen zu anderen Organisationen oder Akteurinnen oder Akteure können Stiftungen ihr Angebot erweitern oder ihre Interessen kollektiv vertreten.

Die Ziele dieser Evaluation beziehen sich vornehmlich darauf, die Stiftungen bzw. GmS konkret bei der Angebotsgestaltung und Programmzielsetzung und -erreichung zu unterstützen, um zu einer evidenzbasierten Verbesserung der Angebotsqualität an Ganztagschulen beizutragen.

2.6 Lehrerüberzeugungen und -handeln:

Stiftungen sind somit diejenigen Akteurinnen, die die Programmziele sowie das Fachpersonal der außerschulischen Ganztagsangebote selektieren, ggf. instruieren sowie einen Teil der Rahmenbedingungen vor Ort vorgeben. Da das Fachpersonal somit die direkte Schnittstelle zwischen Programm/Stiftung und Unterricht darstellt, lässt sich argumentieren, dass es demnach ein Vehikel der Realisierung der Stiftungsziele darstellt. Somit stellt sich die

Frage, welche Faktoren dazu führen, welche Inhalte und welche Methoden das Fachpersonal einsetzt, um den Unterricht und konkreter die Lern- und Erwerbsgelegenheiten der Schülerinnen und Schüler zu gestalten.

In diesem Kontext sollen subjektive (Lehrer-)Überzeugungen als ein möglicher Einflussfaktor herangezogen werden, der in dieser Evaluation genauer untersucht werden soll. Der Begriff der (Lehrer-)Überzeugungen entstammt aus der Lehrerforschung, wobei sich im Laufe der Zeit verschiedene Definitionsstränge entwickelten (vgl. Drexel, 2013). Dies erschwert zum Teil die Passung und Adäquanz mit den Überzeugungen des pädagogischen Fachpersonals. Allerdings erscheint dies – aufgrund der fehlenden Studienlage – eine sinnvolle Herangehensweise, um unterrichtswirksame Überzeugungen des Fachpersonals identifizieren zu können. Demzufolge sollen folglich lediglich jene Definitionen dargestellt werden, die sich nach Einschätzung des Autors dieser Studie ebenfalls auf das Fachpersonal übertragen lässt. In dieser Arbeit wird der Terminus *Überzeugung* nach den nun folgenden Definitionssträngen verwendet, wobei kritisch anzumerken ist, dass es weitere Begrifflichkeiten gibt, die sich in diesem Kontext und teilweise mit derselben Bedeutung anwenden ließen (vgl. Drexel, 2013). So konstatiert Voss et al. (2011), dass sich der Terminus *(Lehrer-)Überzeugung* nicht eindeutig von weiteren Begriffen wie Subjektiven Theorien, Lehrereinstellungen etc. abgrenzen ließe. Demzufolge soll in dieser Arbeit der Begriff *Überzeugung* synonym mit den Begriffen Subjektive Theorie sowie (Lehrer-)Einstellung verwendet werden, um jeweils andere Begriffsaspekte zu betonen, wobei sie sich semantisch auf die nun folgenden Definitionen von (Lehrer-)Überzeugungen beziehen:

Zunächst beschrieb Pajares (1992) (Lehrer-)Überzeugungen (teachers beliefs) als ein kognitives System: Dieses wird einerseits durch persönliche Erfahrungen – insbesondere aufgrund der eigenen Schulbiographie – früh in der Lebensbiographie geformt und von kulturellen und sozialen Erlebnissen geprägt. Überzeugungen werden von Pajares (1992) als zeitinvariant beschrieben, die auch inhaltlich Gegensätze beinhalten können, ohne dass sich das gesamte Überzeugungssystem ändert. Darüber hinaus beschreibt Pajares (1992) Überzeugungen als verhaltensbeeinflussende Faktoren, die somit insbesondere das Lehrerhandeln und somit die Unterrichtsgestaltung prägen. Drexel (2013) beschreibt, dass die (Lehrer-)Überzeugungen einerseits das Handeln beeinflussen und andererseits, dass das (Lehrer-)Handeln die (Lehrer-)Überzeugungen beeinflusst. Somit kann konkludiert werden, dass diejenigen Erfahrungen, die das Fachpersonal bei der Konzeption und Realisierung der Ganztagsangebote macht, wiederum die eigenen Überzeugungen prägt. Somit lassen sich durch die Vorgabe konkreter Rahmenbedingungen seitens der Stiftungen und der Schule bestimmte Situationen kreieren, die wiederum zu bestimmten Erfahrungen führen könnten. Dies ließe sich als ein weiteres Steuerungsinstrument beschreiben, um die (Lehrer-)Überzeugungen des Fachpersonals zu beeinflussen. Des Weiteren definieren Baumert et al. (2011) Überzeugungen als Teil ihres Kompetenzmodells: Überzeugungen beinhalten somit das Fähigkeitsselbstkonzept, subjektive gegenstandsspezifische Lern- und Lehrtheorien sowie epistemologische Überzeugungen. Letztlich soll noch die Definition von Kleickmann et al. (2010) angeführt werden, welcher (fachspezifische) Überzeugungen beschreibt als Vorstellungen, was „guter“ und „weniger guter“ Unterricht ist sowie welche Lernvoraussetzungen die Schülerinnen und Schüler für ein bestimmtes Fach benötigen.

Ausgehend von dieser Definitionen können nun zweierlei Fragen gestellt werden: Einerseits, inwiefern diese (Lehrer-)Überzeugungen durch Stiftungen identifiziert werden (können) und inwiefern Stiftungen Einfluss auf das Lehrerhandeln haben können, in

Anbetracht der curricularen und methodisch-didaktischen Freiheiten des Fachpersonals bei der Unterrichtskonzeption. Andererseits ergibt sich daraus die Frage, inwiefern jene (Lehrer-)Überzeugungen veränderbar sind. Drexel (2013) begründet die Stabilität der (Lehrer-)Überzeugungen darin, dass neue Lerninhalte kognitiv gefiltert werden, abhängig des bestehenden kognitiven Systems. Veränderungen bedürfen somit einer tiefenorientierten kognitiven Verarbeitung der Lerninhalte, welche wiederholt dargeboten werden; mit dem Ziel, die bestehenden Überzeugungen infrage zu stellen (vgl. Voss et al., 2011).

2.7 Kooperation zwischen Ganzttag und externen Kooperationspartnern

Ausgehend von der bisherigen Argumentation und theoretischen Hinführung soll nun überlegt und dargestellt werden, wie die Studienlandschaft die Kooperation zwischen Ganzttag und externen Kooperationspartnern beschreibt: In StEG wurden Ganzttagsschulen befragt, welche Probleme und Schwierigkeiten auftreten, neue Angebote in den Ganzttag zu implementieren (Jugendinstitut, 2019): Die Ganzttagsschulen beschrieben den Prozess als zeitintensiv, welcher gleichzeitig dem Durchhaltevermögen seitens der Akteurinnen und Akteure bedarf. Dies betrifft die Implementation des Angebots, aber auch alle damit verbundenen Prozesse, insbesondere der daraus entstehende Bedarf der multiprofessionellen Kooperation (Jugendinstitut, 2019). Des Weiteren stellen sich die von StEG angestrebten Zieldimensionen der Kooperation im Ganzttag – ein „Wir“-Gefühl aller pädagogisch Tätigen sowie die gleichberechtigte Partizipation – ebenfalls erst nach langer Zeit beidseitiger Bemühungen ein (Jugendinstitut, 2019).

Zentral für die Zusammenarbeit ist demgemäß die Möglichkeit zur Partizipation (Jugendinstitut, 2019): Dadurch wird den pädagogischen Akteurinnen und Akteuren vermittelt, dass ihre Kompetenz wahrgenommen wird. Allerdings wirken sich stark hierarchische Organisationsformen, die insbesondere hierarchisch zwischen schulischen und außerschulischen Akteurinnen und Akteuren trennen, hemmend auf die Kooperation aus (Jugendinstitut, 2019). StEG schlägt eine transparente Kommunikation zwischen den schulischen und außerschulischen Akteurinnen und Akteuren vor, welche alle Beteiligte einbezieht (Jugendinstitut, 2019). Das bedeutet, dass alle Beteiligte gleichermaßen die Möglichkeit besitzen, an den Informationen und Entscheidungen des pädagogischen Alltags teilzuhaben (Jugendinstitut, 2019). Nach StEG entstehen Kooperationsstrukturen im Diskurs, wobei Differenzen zwischen den Kooperationspartnern zu unterschweligen Konflikten führen können, welche sich wiederum nachteilig auf die Kooperation auswirkt (Jugendinstitut, 2019). Hierbei schlagen die Autorinnen und Autoren von StEG vor, die gemeinsame Beziehungsarbeit offen zu thematisieren und gegebenenfalls durch Teamsupervision oder Teamentwicklung zu fördern (Jugendinstitut, 2019). Allerdings benötigen diese Prozesse Zeit, insbesondere das Aufbrechen langjährig gewachsener Strukturen, die zwischen Lehrkräften und dem weiteren pädagogischen Personal trennen (Jugendinstitut, 2019). Gleichzeitig stellen diese Austausch- und Kommunikationsprozesse gute Anlässe für Schulentwicklungsprozesse dar, wonach insbesondere die Schule davon profitieren kann (Jugendinstitut, 2019). StEG beschreibt, dass die Kooperation zwischen dem pädagogischen Fachpersonal und den schulischen Akteurinnen und Akteuren eine wichtige Voraussetzung zur Umsetzung von individueller Förderung ist (Jugendinstitut, 2019). Gleichzeitig ermöglicht die Kooperation mit externen Kooperationspartnern das Schaffen von neuen Methoden, um den Ganzttag weiterzuentwickeln und auf die Schülerbedarfe einzugehen (Jugendinstitut,

2019). Dies setzt des Weiteren passende räumliche Strukturen und zeitliche Ressourcen voraus (Jugendinstitut, 2019).

Der Aspekt der Kooperation lässt sich auch auf einer Zeitdimension betrachten: So beschreiben Van Santen und Seckinger (2003) die Dauer der Kooperation als ein weiteres Qualitätsmerkmal der Kooperation und skizzieren drei Phasen, welche Kooperationsbeziehungen über die Zeit durchlaufen und welche mit bestimmten Qualitätsmerkmalen verbunden sind: Zunächst die Gründungsphase, in welcher sich Schule und externer Kooperationspartner programm- und angebotsbezogen konzeptionell definieren und erste Annäherungsversuche beschreiten. Anschließend die Etablierungsphase, in welcher sich die Kooperation als struktureller Bestandteil etabliert und abschließend die Konsolidierungsphase, in welcher sich die strukturellen Kooperationsprozesse ausdifferenzieren. Analog dazu beschreibt Thielen (2010) die Zeit, die eine Kooperation bereits besteht, als einen der zentralsten Qualitätsfaktoren.

2.8 Evaluation

Ausgehend von den dargestellten Rahmenbedingungen soll im Folgenden festgelegt werden, was unter Evaluation zu verstehen ist und welche Funktion Evaluationen in dem Kontext von Ganztagsangeboten einnehmen können. Für wissenschaftliche Evaluationen existiert eine große Bandbreite an Definitionsansätzen, die jeweils andere Aspekte von Evaluation beschreiben oder betonen. Folglich sollen nun einige ausgewählte Definitionsansätze dargestellt werden, die sich nach Einschätzung des Autors dieser Arbeit auf den Bereich der Evaluierung von Ganztagsangeboten übertragen lassen und gleichzeitig individuelle Schwerpunkte setzen, nach denen sich diese Arbeit ausrichtet. So soll zunächst allgemein die Definition von von Kardorff und Schönberger (2020) herangezogen werden, um den Evaluationsbegriff allgemein zu umschreiben: Diese beschreiben Evaluation als Prozess und Produkt einer Bewertung, welche sich auf einen spezifischen Gegenstand bezieht. Von Kardorff und Schönberger (2020) definieren vor diesem Hintergrund drei Facetten von Bewertungen, auf die sich ihr Evaluationsbegriff bezieht:

1. Bewertungen beziehen sich einerseits auf **Ist-Soll-Vergleiche** zwischen dem Gegenstand auf der einen und den angestrebten Veränderungen oder Normen auf der anderen Seite (Kardorff & Schönberger, 2020). Hierdurch wird es möglich, im Zuge der Evaluation den Evaluationsgegenstand anhand eines normativen Qualitätsmodells im Sinne von Ist-Soll-Vergleichen zu bewerten. Dies steht in Einklang mit den Beobachtungen von Burkard und Eikenbusch (2005), nach denen sich Evaluationen immer stärker gegenständlich an normativ-komparativen Qualitätsbegriffen orientieren.
2. Des Weiteren können Bewertungen **Vorher-Nachher-Vergleiche** auf einer Dimension beschreiben wie z. B. Verbesserungen des Leseverständnisses oder der motivationalen Orientierung (Kardorff & Schönberger, 2020). Dies deckt vor diesem Hintergrund eine wichtige Funktion von Evaluationen ab, nämlich der Aussage, ob sich durch eine pädagogische Intervention (statistisch signifikante) Dispositionsveränderungen beobachten lassen, die sich auf die pädagogische Intervention zurückführen lassen.
3. Abschließend können Bewertungen als **Vergleiche zwischen konkurrierenden Personen, Angeboten, Objekten, Programmen oder Zuständen** bezeichnet

werden. Aufgrund der großen Angebotsvielfalt und der mangelnden reglementierten Standardisierung von Ganztagsangeboten und darüber hinaus des großen Einflusses des Kontextes stellt sich die Frage, inwiefern Vergleiche zwischen unterschiedlichen Programmen sinnvoll sind. Es lässt sich vorstellen, dass hierbei die angebotsrelevanten Faktoren sehr unterschiedlich ausgeprägt sein können und somit einen Vergleich erschweren (z. B. Programmziele, Programmmittel, Programmpersonal, Einbettung des Programms in Schule, Schüler- und Elternschaft, Kompetenzen der Schule im Umgang mit externen Kooperationspartnern usw.). In dieser Arbeit soll deshalb und aufgrund der lückenhaften Studienlandschaft zur Angebotsqualität auf den Vergleich mit anderen Programmen verzichtet werden. Es sollen lediglich durch andere Evaluationen und Studien beschriebene Herausforderungen und Wirkzusammenhänge zur Interpretation der Ergebnisse dieser Arbeit herangezogen werden.

Welche genauen definitorischen Eigenschaften ein Evaluationsgegenstand haben soll, wird in der hier zitierten Literatur allerdings nicht expliziert. Implizit werden allerdings konkrete sowie abstrakte Beispiele von Evaluationsgegenständen beschrieben, ausgehend von Objekten über Prozesse und Zustände (vgl. Kardorff & Schönberger, 2020). Durch diesen unspezifischen und breit gefächerten Gegenstandsbezug lässt sich beinahe alles evaluieren, wobei vorstellbar ist, dass insbesondere pragmatische Machbarkeitsüberlegungen und die Operationalisierbarkeit der theoretischen Konstrukte diesen Gegenstandsbereich einzugrenzen vermögen. Des Weiteren erscheint es notwendig, dass eine Passung zwischen den zum Vergleich herangezogenen normativen (Qualitäts-)Kriterien und dem Evaluationsgegenstand besteht, insofern als das normative Modell den Gegenstand, dessen Eigenschaften und dessen Eigenschaftsausprägungen realistisch beschreibt.

Zuzüglich zu der dargestellten Definition von Kardorff und Schönberger (2020) sollen noch zwei weitere Definitionsansätze dargestellt werden, die bestimmte Facetten dieser Arbeit genauer definieren. Zunächst definiert Giel (2013) Evaluation als ein Qualitätsverbesserungsinstrument. Das bedeutet, Evaluation kommt selbst eine bestimmte Wirkung auf den Evaluationsgegenstand zu, nämlich die der Qualitätsverbesserung. Hieran anknüpfend lassen sich die Erkenntnisse von Helmke (2003) anführen: Dieser beschreibt durch Evaluation ausgelöste Veränderungsprozesse als lang und störanfällig, wobei wichtige Gelingensbedingungen die Merkmale des innerschulischen Kontextes sowie die personalen Merkmale der Beteiligten darstellen. Abschließend wird die Definition von Rogers (2002) herangezogen, welcher die Veränderung und Verbesserung der Planung und Umsetzung von Programmen als zentrale Aufgabe von (Programm-)Evaluationen beschreibt. Nach dieser Definition lässt sich diese Arbeit auch selbst als Intervention einstufen, die Planungs- und Realisierungsprozesse des Programms verändern bzw. verbessern kann oder – normativ gesprochen – auch soll.

Mit Hense und Mandl (2003) lässt sich der Evaluationsbegriff noch weiter spezifizieren, nämlich nach den Funktionen, die Evaluationen einnehmen können:

1. **Entwicklungsfunktion:** Die Evaluation dient primär dem Gewinnen von Erkenntnissen zur konkreten Verbesserung des Evaluationsgegenstandes.
2. **Entscheidungsfunktion:** Die Evaluationsergebnisse werden vornehmlich als Entscheidungsgrundlage (von den Auftraggebenden) herangezogen.

3. **Kontrollfunktion:** Die Evaluation dient primär der Kontrolle des Evaluationsgegenstands seitens der Auftraggebenden.
4. **Lernfunktion:** Die Evaluation dient hauptsächlich dem Gewinnen von Erkenntnissen über den Evaluationsgegenstand.

Diese Evaluationsfunktionen müssen sich allerdings nicht ausschließen und können auch gleichzeitig – mehr oder minder stark ausgeprägt - auftreten (Hense & Mandl, 2003). Evaluationen lassen sich darüber hinaus kategorial in summative und formative Evaluationen einteilen (Kardorff & Schönberger, 2020):

1. **Summative Evaluationen** beschreiben das Messen (und Bewerten) des Outputs oder Outcomes von bestimmten Prozessen (z. B. Programme). Hierzu werden größtenteils quantitative Methoden herangezogen, die konkrete quantitative Vergleiche anhand vorgefundener Zusammenhänge oder Unterschiede aufzeigen. Im Zuge dessen soll ein wertendes Urteil über den Evaluationsgegenstand gefällt werden.
2. **Formative Evaluationen** beziehen sich auf die Begleitung und evtl. die Steuerung des evaluationsgegenständlichen Prozesses. Hierbei gehen die Autoren von einer wechselseitigen Beeinflussung zwischen Erkenntnisgegenstand (hier: Programm) und Forschenden aus, was zu Programm- oder Programmzieländerungen führen kann.

Diese Arbeit vereint sowohl summative als auch formative Evaluationsaspekte, insbesondere mit dem Ziel, einen Erkenntnisgewinn für ENB zu generieren.

Giel (2011) beschreibt drei Formen von Programmevaluationen: Externe, interne sowie Selbstevaluation:

1. Bei der **externen Evaluation** ist der Evaluierende nicht Teil der Organisation, das bedeutet, er besitzt keine weitere Funktion im Programm, außer die des Evaluierenden.
2. Bei der **internen Evaluation** stammt der Evaluierende aus der Organisation.
3. Bei der **Selbstevaluation** untersuchen die Programmdurchführenden ihre eigenen Strategien und Interventionen systematisch, ob diese funktionieren oder nützlich sind.

Bei dieser Arbeit handelt es sich um eine externe Evaluation. Des Weiteren beschreiben Kardorff und Schönberger (2020), dass die Evaluationsforschung häufig auf alle möglichen sozialwissenschaftlichen Methoden zurückgreift, um den Evaluationsgegenstand zu untersuchen. Das bedeutet, dass Evaluation nicht auf bestimmte Methoden festgelegt ist, sondern sich methodisch auf ihren Gegenstand und die gegebenen Rahmenbedingungen ausrichtet.

Im Rahmen einer Evaluation wird gewöhnlicherweise ein Evaluationsmodell erstellt, welches den Rahmen und den Ablauf der Evaluation festlegt. In dem Evaluationsmodell werden insbesondere die selektiven Schwerpunkte der Evaluation festgelegt, nach denen sie den Evaluationsgegenstand untersucht. Diese Arbeit orientiert sich an dem CIPP-Modell von Stufflebeam (1971), welches Evaluationen in vier Bereiche gliedert und im Folgenden auf den Bereich der Programmevaluation übertragen dargestellt wird:

1. **Kontext-Evaluation:** Identifikation der Entstehungsgründe eines Programms.
2. **Input-Evaluation:** Identifikation und Analyse der einem Programm zur Verfügung stehenden Ressourcen.
3. **Prozess-Evaluation:** Analyse der Programmdurchführung.

4. **Produkt (Output, Outcome):** Identifikation und Analyse der kurzfristigen (Output) und/oder langfristigen (Outcome) Effekte eines Programms sowie deren Bewertung.

Fischer et al. (2011) erweiterten dieses CIPP-Modell im Rahmen der StEG-Studie, indem sie jedem Bereich verschiedene, ganztagschulspezifische Unterkategorien zuordneten: *Kontext und Input* wird zu einer Kategorie zusammengefasst und durch schulische Rahmenbedingungen, externer und interner Kontext sowie familiärer Kontext erweitert. Der *Prozess* wird aufgeteilt in Angebot und Nutzung, wobei das Angebot Prozessmerkmale des Unterrichts und des Angebots beschreibt sowie die Schulqualität und die Nutzung sich auf den Unterricht und die Angebote bezieht. Der *Outcome* wird in fachübergreifenden Wirkungen sowie Wirkungen auf die Leistungen und Schulerfolg beschrieben. All diesen Prozessen wird die Schulentwicklungsarbeit zugeordnet, das bedeutet, dass durch jede Evaluationsform Schulentwicklungsarbeit geleistet.

Für Evaluationen werden gemeinhin bestimmte Qualitätsstandards definiert, die es gilt, möglichst einzuhalten. Diese Standards wurden durch die Deutsche Gesellschaft für Evaluation definiert und umfassen vier Oberkategorien (DeGEval, 2024):

1. **Nützlichkeit:** Die Evaluation besitzt eine bestimmte Funktion, welche konkret benannt wird.
2. **Durchführbarkeit:** Die Evaluation umfasst angemessene Verfahren, sodass einerseits die Evaluationsziele entsprechend umgesetzt werden und andererseits nicht zu viel Belastung für die Beteiligten entstehen.
3. **Fairness:** Die Evaluation orientiert sich an ethischen Richtlinien.
4. **Genauigkeit:** Der Evaluationsgegenstand wird genau beschrieben und empirisch unter Einhaltung der anerkannten Gütekriterien untersucht.

Diese Arbeit hat sich stark an diesen Standards orientiert, wobei das größte Augenmerk auf den ethischen Standards bzgl. des Umgangs mit persönlichen Daten der Beteiligten liegt.

3. Evaluation des Programms „Eine Note besser!“

3.1 Beschreibung des Programms

Im Folgenden wird das Programm „Eine Note besser!“ (ENB) vorgestellt werden, welches den Evaluationsgegenstand darstellt. Hierbei sollen lediglich ausgewählte Aspekte dargestellt werden, da weiterführende Informationen auf dem zugehörigen Webauftritt seitens des Programms vermittelt werden. Die folgenden Informationen entstammen jenem Webauftritt. Ziel des Programms ist nach eigenen Aussagen der Programmverantwortlichen das Erhöhen der Chancengerechtigkeit für Kinder mit schwierigen Startbedingungen, indem man gezielt Förderangebote für Schülerinnen und Schüler des Ganztags an Grundschulen in Stadtteilen mit niedrigem Sozialindex anbietet. Um diese Ziele zu erreichen, formuliert ENB fünf Qualitätsmerkmale, an denen es sich orientieren will:

1. **Kontinuität:** Die Kinder werden über die gesamte Grundschullaufbahn durch dieselben Personen gefördert, um eine Bindung mit diesen aufzubauen.
2. **Vielfältigkeit:** Die Kinder sollen eine umfassende Förderung erfahren, die sprachliche, sozial-emotionale, fachliche und kreative Inhalte umfasst.
3. **Adaptivität:** Das Programm passt sich stets an die sich verändernden Lernvoraussetzungen der Kinder an.

4. **Multiprofessionalität:** Die Fachkräfte, die das Programm durchführen, sollen verschiedene professionelle Hintergründe aus dem Bildungsbereich vereinen und miteinander kooperierend die Kinder fördern.
5. **Wirkungsorientierung:** Das Programm ist darauf ausgelegt, Wirkungen zu erzielen und den Stand der Zielerreichung zu messen, um auf dieser Grundlage ggf. Anpassungen zu treffen.

Das Projekt entstand 2014 auf Initiative der Stiftungen Castringius Kinder & Jugend Stiftung München und der Stiftung Gesellschaft macht Schule (GmS) an einer Münchner Grundschule und wurde vier Jahre lang durch die LMU München begleitet. In diesem Zeitraum wurde eine Grundschulklasse mit Förderangeboten unterstützt. Nach Abschluss des Projekts wurden die Erkenntnisse zu einem konkreten Programmkonzept und einer Wirkungslogik integriert und an einer neuen Grundschule umgesetzt. Hierbei handelt es sich um die Grundschule an der Walliser Straße in München. Der Kontakt zu der Schule wurde durch die Programmleitung hergestellt und erfüllt nach Aussagen der Programmverantwortlichen die durch den Sozialindex bestimmten Förderkriterien.

Um die Programmziele zu erreichen, wurden vier Förderbereiche bestimmt, die jeweils in einer bis zwei Schulstunden realisiert werden:

1. **Sprachliche Projekte:** Förderung der sprachlichen und Lesekompetenzen.
2. **Sozio-emotionale Projekte:** Förderung der sozio-emotionalen Kompetenzen.
3. **Kreative Projekte:** Förderung der Kreativität der Kinder.
4. **Individuelles Coaching:** Unterstützung und Förderung der Kinder in Hinsicht auf ihre Lernvoraussetzungen.

Des Weiteren unterstützt das Programm einmal die Woche die Lehrkraft im Unterricht bei der Betreuung und Unterstützung der Aufgabenbearbeitung im Rahmen des Wochenplans. Allerdings hat sich über den Programmverlauf die Personalanzahl verändert, wobei lediglich neues Personal hinzugekommen ist. Grundsätzlich gibt es dabei eine Programmleitung, die u. a. für die Organisation und die Planung des Programms zuständig ist. Daneben gibt es noch für jedes der beschriebenen Projekte eine Fachkraft, die unterschiedliche professionelle Hintergründe besitzen. Hierunter befinden sich Professionen wie Theaterpädagogik, Fachpersonal für Leseförderung sowie Soziale Arbeit. Das Programm wird zu großen Teilen von der Castringius Stiftung sowie GmS finanziert.

Das Programm hat – ausgehend von den Erkenntnissen des ersten Projekts – eine Wirkungslogik definiert, die angelehnt an die Programmtheorie der Theoriebasierten Evaluation die angestrebten Wirkmechanismen beschreiben soll, auf die das Programm abzielt. Diese werden in Abbildung 2 dargestellt und orientieren sich zum größten Teil an Phineo – Kursbuch Wirkung (Kurz et al., 2013), welches Programmen Unterstützung bei der Ausformulierung solcher Wirkungslogiken geben.

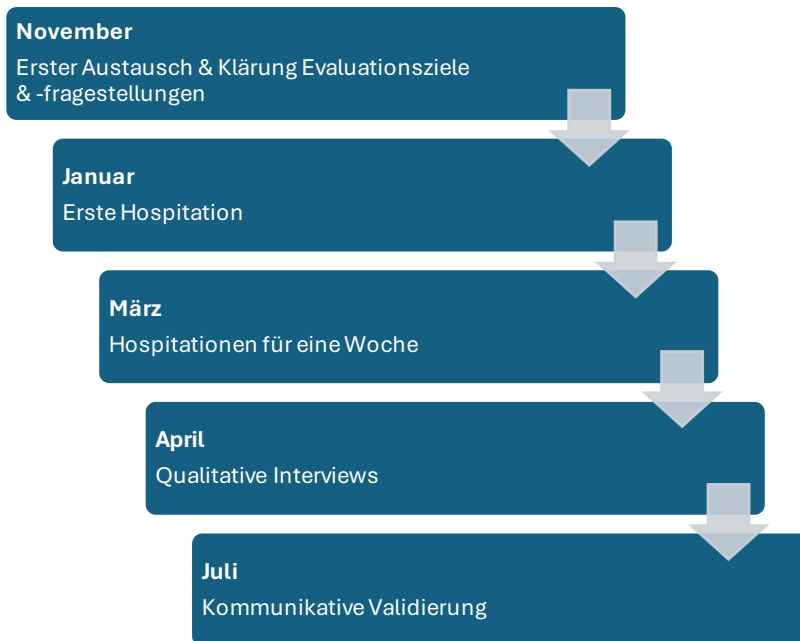
Abbildung 2 Wirkungslogik von ENB

Zielgruppen	Leistungen	Erwartete Wirkung
A. Schulkinder im Ganzttag (1. bis 4. Jgst.)	<p>Wöchentliche Kurse im Ganzttag mit beständiger Bezugsperson</p> <p>* Projektarbeit in der halben Klasse: <u>Sozial-emotional</u> z.B. Sozialkompetenztraining, Persönlichkeitsstärkung <u>Sprachlich</u> z.B. Schreib-/Sprachwerkstatt, Hörspiel-Projekt <u>Kreativ, motorisch</u> z.B. kulturpädagogische Angebote</p> <p>* Individuelles Coaching (Einzel & in Kleingruppen, nach Bedarf): - Sprachförderung - Deutsch- und Matheförderung - Sozial-emotionales Coaching</p>	<p>Die teilnehmenden Kinder verbessern ...</p> <p><u>Lebenslage</u></p> <ul style="list-style-type: none"> - ... ihre Lern- und Leistungsmotivation - ... ihre Chancen auf einen Übergang auf eine weiterführende Schule, der ihrem Potenzial entspricht <p><u>Handeln</u></p> <ul style="list-style-type: none"> - ... ihr Lern- und Arbeitsverhalten <p><u>Wissen und Fertigkeiten</u></p> <ul style="list-style-type: none"> - ... ihre personalen und sozialen Fähigkeiten (u. a. Kommunikations-, Selbstbehauptungsfähigkeit) - ... ihre Sprach- und Lesekompetenz sowie ihre fachspezifischen Lernfortschritte (u. a. Lernplanziele in Deutsch, Mathematik und Heimat- und Sachunterricht)
B. Eltern	<p>* Elternarbeit (Unterstützung durch die Programm-beteiligten)</p> <p>z.B.:</p> <ul style="list-style-type: none"> - Elternabende - Elterngespräche - Elternbriefe - Elternfeste 	<p>Die Eltern der teilnehmenden Kinder ...</p> <ul style="list-style-type: none"> - ... partizipieren an schulischen Veranstaltungen - ... haben ein beständiges und vertrauensvolles Verhältnis zu den Programmbeteiligten - ... werden bei der Einschätzung der Potentiale ihrer Kinder sowie bei Bildungsentscheidungen unterstützt
C. Lehrkraft	<p>* Beobachtung und kontinuierlicher Austausch der Programmbeteiligten</p> <p>* Elternarbeit</p> <p>* Projektarbeit in der halben Klasse</p> <p>* Individuelles Coaching (Einzel & in Kleingruppen, nach Bedarf)</p>	<p>Die Lehrkraft der teilnehmenden Klasse wird entlastet durch...</p> <ul style="list-style-type: none"> - ... die Unterstützung bei der Umsetzung des Lehrplans - ... die Unterstützung bei der Aufdeckung und Lösung sozial-emotionaler Probleme der Kinder - ... die Einbindung in ein multiprofessionelles Team

3.2 Chronologischer Verlauf der Evaluation

Im Folgenden soll der Verlauf und die zentralen Meilensteine des Evaluationsprozesses dargestellt werden. In Abbildung 3 werden diese chronologisch nach Monaten aufgelistet. Im November fand der erste Austausch zwischen dem Autor dieser Arbeit und den Programmbeteiligten statt. Hierbei wurden die Evaluationsziele und -fragenstellungen ausgehandelt, die zur Verfügung stehenden Daten sowie der Umfang der qualitativen Interviews. Im Januar hospitierte der Autor dieser Arbeit einen Tag und im März für eine Woche an der Schule. Im April wurden die qualitativen Interviews realisiert und im Juli wurden die Ergebnisse aufbereitet, den Programmbeteiligten vorgestellt sowie kommunikativ validiert, indem einerseits die eigenen Interpretationen dargestellt und dem Diskurs ausgesetzt und andererseits Interpretationsansätze der Beteiligten geäußert wurden. Neben diesen dargestellten Meilensteinen gab es des Weiteren einen regen Email- und telefonischen Austausch, insbesondere zwischen der Programmleitung und dem Autor dieser Studie.

Abbildung 3 Chronologischer Verlauf und Meilensteine des Evaluationsprozesses



3.3 Evaluationsziele, Evaluationsmodell & Fragestellungen

Die Evaluationsziele dieser Arbeit wurden recht unspezifisch formuliert und dem Autor dieser Arbeit wurden große Freiheiten gewährt, eigene Schwerpunkte zu setzen. Lediglich die zur Verfügung stehenden Daten und der zeitliche Rahmen lassen sich als externe limitierende Faktoren beschreiben. Im bereits beschriebenen ersten Austauschtreffen zwischen dem Autor dieser Studie und den Programmverantwortlichen wurde festgelegt, dass die Evaluation einerseits Wirksamkeitsaussagen über das Programm treffen sollte und andererseits Erkenntnisse generieren soll, welche sich für zukünftige, ähnlich ausgerichtete Programme wie ENB übertragen ließen. Dies hat den Grund, dass das Programm ENB nach Beendigung der Förderphase der in dieser Arbeit beschriebenen Grundschulklasse nicht mehr auf diese Weise aufgesetzt wird, sondern die dort gewonnenen Erkenntnisse zur Verbesserung und Modifikation der weiteren bestehenden Programmen und Schwerpunkten von ENB herangezogen werden sollen.

Um die zur Verfügung stehenden Ressourcen sowie die Evaluationsfragestellungen festzulegen, fand wie in Kapitel 3.2 dargestellt ein Treffen zwischen dem Autor dieser Studie, einer Begleitperson des Autors, der Programmleitung und der Leitung der Stiftung (GmS) statt. Im Vorfeld des Treffens hatte der Autor dieser Studie bereits die zur Verfügung stehende Literatur seitens der Stiftung gesichtet und eigene interessierende, thematische Schwerpunkte ausgewählt und zusammengestellt, die sich zur Evaluation eignen könnten. Anzumerken ist, dass dieser Auswahlprozess aufgrund der Kurzfristigkeit nicht durch die theoretische und empirische Studienlage fundiert wurde, sondern mittels logischer Überlegungen und individueller Präferenzen seitens des Autors gelenkt wurde. Der Aushandlungsprozess, der während dieses Treffens stattfand, bezog sich somit auf die Ausarbeitung des Autors und die Passung mit den Zielen und Überlegungen seitens der Programmverantwortlichen. Resultat waren jene Inhalte, die der Autor in Anhang (Zu

erhebende Konstrukte) nach dem Treffen festhielt und in dieser Form den Programmverantwortlichen nach dem Treffen zukommen ließ.

Neben den allgemeinen Evaluationsfragestellungen wurden des Weiteren die zur Evaluation zu verwendende Datenbasis sowie die Interviewformate bestimmt. Die Datenbasis bezieht sich auf diejenigen Daten, die zur quantitativen Analyse herangezogen werden könnten. Hierbei wurde sich konsensual auf die Verwendung der Daten der ELFE- und SELLMO-Tests sowie die Schulnoten der Schülerinnen und Schüler geeinigt. Gleichzeitig wurde vereinbart, dass darüber hinaus qualitative Interviews mit dem Fachpersonal und der Lehrkraft durchgeführt werden. Die damit verbundenen Interviewformate beziehen sich demnach auf die Auswahl der Stichprobe sowie die Dauer der Interviews. Unter der Stichprobe befanden sich somit drei pädagogische Fachkräfte, die Programmleitung von ENB sowie die Klassenlehrkraft der geförderten Klasse. Im Laufe des Evaluationsprozesses wurden allerdings die zu Beginn formulierten Fragestellungen inhaltlich weiter spezifiziert und teilweise modifiziert, um eine stärkere Passung mit den Evaluationszielen herzustellen und auf neue Erkenntnisse zu reagieren.

In Tabelle 1 ist das Evaluationsmodell dieser Arbeit abgebildet, welches sich an Fischer et al. (2011) orientiert und konkret auf den Evaluationsgegenstand ENB und dessen Wirkungslogik (Kapitel 3.1) adaptiert wurde. Dieses ist das Produkt sowohl der Aushandlungsprozesse zwischen dem Autor dieser Arbeit und den Programmverantwortlichen als auch den individuellen Schwerpunktsetzungen des Autors dieser Arbeit.

Tabelle 1 Evaluationsmodell dieser Arbeit

Kontext & Input		Prozess		Wirkung/Output	
<i>ENB</i>	Einstellung des Fachpersonals	<i>Angebotsgestaltung</i>	Allgemeiner Ablauf	<i>Leseverständnis</i>	Statistisch signifikante Verbesserungen
	Zielorientierung des Fachpersonals		Angebotsqualität		Statistisch signifikante Verbesserungen
	Professionalität des Fachpersonals	<i>Multiprofessionelle Kooperation</i>	Zwischen Fachpersonal	<i>Noten</i>	Verlauf über die Zeit
Evidenzbasierung des Programms	Mit Schule				
<i>Schule</i>	Rahmenbedingungen für ENB	<i>Einbettung des Programms in Schule</i>	Wahrgenommene Partizipationsmöglichkeiten Aushandlungsprozesse zwischen ENB mit Schule		

Um die Evaluationsziele zu erreichen wurde sich im Austausch mit Herrn Professor Eckert entschieden, die Evaluation in einen summativen und einen formativen Teil zu gliedern, um einerseits ENB und GmS summative und wissenschaftsbasierende Werturteile zur Verfügung zu stellen und andererseits um für zukünftige Programme eine Erkenntnis- und Ideengrundlage darzubieten, die zur Programm- und Angebotsqualitätsverbesserung verwendet werden können. In Tabelle 2 werden die in dieser Arbeit zu untersuchenden Fragestellungen aufgeführt, welche sich aus dem Evaluationsmodell ergaben.

Tabelle 2 Fragestellungen dieser Evaluation

Formative Evaluation	Summative Evaluation
1.1 Welche Verläufe des Leseverständnisses lassen sich über die vier Jahre beobachten?	2.1 Sind Verbesserungen in dem Leseverständnis der Kinder über die vier Jahre beobachtbar?
1.2 Wie gestaltet das Fachpersonal die Unterrichtsstunden und was ist ihm dabei wichtig?	2.2 Gibt es positive Veränderungen in der motivationalen Orientierung der Kinder?
1.3 Welches Verständnis hat das Fachpersonal von den Programmzielen?	
1.4 Wie nimmt das Fachpersonal die Kooperation mit und die Integration in der Schule wahr?	
1.5 Welche Einstellung hat das Fachpersonal gegenüber dem Programm?	
1.6 Welches Verständnis hat das Fachpersonal in Bezug auf seine individuelle Rolle im Programm?	
1.7 Wie trifft das Fachpersonal diagnostische Urteile und Wirksamkeitsurteile?	
1.8 Wie lässt sich das Programm evidenzbasierter gestalten?	
1.9 Welche Einstellung hat das Fachpersonal gegenüber quantitativen diagnostischen Instrumenten und deren Integration in die pädagogische Praxis?	
1.10 Welche Strategien wendet das Fachpersonal im pädagogischen Umgang mit quantitativen diagnostischen Instrumenten an?	

3.4 Konzeption der Evaluation

Bei dieser Arbeit handelt es sich um eine externe Evaluation. Der Autor dieser Arbeit hatte darüber hinaus keine weiteren Beziehungen mit dem Programm und den Programmteilnehmenden. Die Ziele dieser Evaluation ergeben sich hinsichtlich der Beantwortung der im Laufe des Aushandlungsprozesses formulierten Fragestellungen, wobei, wie bereits

beschrieben, die Ziele höherer Ordnung in der Formulierung eines summativen Werturteils und des Zur-Verfügung-Stellens von Erkenntnissen zur Verbesserung des Programms liegen. Vornehmliches Ziel dieser Arbeit war es, sicherzustellen, dass die Evaluationsziele sowie die Fragestellungen in Einklang mit den Bedarfen der Programmbeteiligten sind, weshalb auf einen regen Austausch zur Informationsbeschaffung sowie Rückmeldung Wert gelegt wurde. Dies lässt sich auch im Wortlaut der Evaluationsfunktionen ausdrücken: Diese Evaluation nimmt sowohl eine Entwicklungs- als auch eine Kontrollfunktion ein.

Wie bereits erwähnt, sollen die im Laufe der Evaluation die gewonnenen Erkenntnisse bewertet werden. Dies geschieht in dieser Arbeit vornehmlich anhand normativer Vergleiche und nicht dem Vergleich mit empirischen Daten aus anderen Studien, wie z. B. die Bewertung der Effektgröße im Leseverständnis anhand des Vergleichs mit den Effektgrößen anderer Programme. Dies hat – wie bereits beschrieben – den Hintergrund, dass es vorstellbar ist, dass sich Vergleiche mit anderen pädagogischen Programmen oder Interventionen aufgrund der Unterschiede in den jeweiligen programm- und kontextbezogenen Faktoren kaum sinnvoll ziehen lassen. Hierbei werden zunächst diejenigen (normativen) Qualitätsindikatoren herangezogen, die ENB sich selbst formulierte und nach denen es das Programm konzipierte und realisierte (siehe Kapitel 3.1). Darüber hinaus werden die in Kapitel 2 dargestellten Theorien und Befunde – insbesondere aus StEG – herangezogen, um Ist-Soll-Vergleiche zu ziehen. Die quantitativen Daten werden insbesondere auch für Vorher-Nachher-Vergleiche herangezogen, um zu Werturteilen hinsichtlich der Wirkungslogik (siehe Kapitel 3.1) zu kommen. Des Weiteren zieht der Autor dieser Studie eigene Bewertungskriterien heran, die sowohl biographisch und in der Sozialisation in bildungsbezogenen und akademischen Kontexten als auch im Austausch mit den Programmbeteiligten sowie in den vom Autor dieser Studie realisierten Hospitationen erworben wurden.

Wie bereits erwähnt, handelt es sich bei dem Evaluationsgegenstand dieser Arbeit um das Programm ENB. Aus den beschriebenen Aushandlungsprozessen, den zur Verfügung gestellten Daten und Erhebungsmöglichkeiten und den Evaluationszielen ergibt sich der Rückgriff auf ein Mixed-Method-Design, bei welchem der Evaluationsgegenstand mittels quantitativer und qualitativer Methoden analysiert wird (Creswell, 2017): Hierdurch wird es möglich, den ausgehandelten summativen und formativen Zieldimensionen dieser Arbeit stärker gerecht zu werden, da einerseits vornehmlich quantitative Erkenntnisse zur summativen und qualitative Erkenntnisse zur formativen Evaluation verwendet werden können. Dies schließt allerdings nicht aus, dass auf Grundlage der qualitativen Erkenntnisse keine summativen Aussagen getroffen werden können, diese werden in dieser Arbeit lediglich nicht quantifiziert geäußert vice versa. Des Weiteren erlaubt ein Mixed-Method-Design die Methodentriangulation in Bezug auf den Forschungsgegenstand, wodurch die verschiedenen Zugänge auf ENB, die im Evaluationsmodell in Kapitel 3.3 dargestellt werden, erst ermöglicht werden (Creswell, 2017).

Im Folgenden soll die Wahl des Mixed-Method-Designs ausführlicher begründet werden: Aufgrund der vorhandenen quantitativen Längsschnittdaten ist es angezeigt, diese auch quantitativ zu analysieren und hiermit die Programmziele summativ zu überprüfen. Da GmS und ENB ebenfalls ermöglichten, Interviews mit dem Fachpersonal sowie der Lehrkraft durchzuführen, musste dahingehend entschieden werden, ob ein quantitativer oder ein qualitativer Zugang den Evaluationszielen stärker gerecht wird. Gegen einen quantitativen Zugang spräche zunächst die sehr kleine Stichprobengröße von fünf Personen (drei Fachkräfte, eine Programmleitung und eine Lehrkraft) sowie fehlende Daten hinsichtlich des

Kontextes und der Schülerschaft, die zur Verfügung stünden. Für einen qualitativen Zugang spräche hingegen einerseits die große Offenheit bei der Untersuchung des Evaluationsgegenstandes – insbesondere vor dem Hintergrund, dass noch wenig Erkenntnisse hinsichtlich der Wirkfaktoren außerschulischer Angebote vorliegen – und andererseits die Möglichkeit der Programmtheorierekonstruktion im Rahmen der Theoriebasierten Evaluation. Ausgehend von dieser Argumentation wurde sich für ein Mixed-Method-Design entschieden.

4. Quantitative Analyse

4.1 Fragestellungen & Hypothesen

Wie bereits beschrieben, wurde im Rahmen des Mixed-Method-Designs quantitative Analysen realisiert, um einen Teil der Fragestellungen zu beantworten. Im Folgenden sollen die durch die quantitative Analyse zu beantwortenden Fragestellungen aus Kapitel 3.3 dargestellt werden. Zu jeder Fragestellung wird anschließend mindestens eine Hypothese formuliert, die anhand der vorhandenen empirischen Daten statistisch überprüft werden.

In Tabelle 3 werden die Fragestellungen und die jeweiligen Hypothesen abgebildet. Hierbei wurde zwischen Fragestellungen der summativen und der formativen Evaluation unterschieden (siehe Kapitel 3). Die Fragestellungen der summativen Evaluation umfassen jeweils mehrere Hypothesen, die es gilt, mittels inferenzstatistischer Methoden an den erhobenen Daten zu überprüfen, um zu einem abschließendem Urteil zu kommen. Die Fragestellungen der formativen Evaluation hingegen sind eher explorativer Natur. Ihr Ziel ist es, die Fragestellungen anhand der Strukturierung, der deskriptiven Darstellung sowie der teilweisen Interpretation der Daten zu beantworten oder zumindest zu approximieren. Die Richtung der Hypothesen der summativen Evaluation ergeben sich aus der Wirkungslogik von ENB in Kapitel 3.1. Da in der Wirkungslogik von ENB beschrieben wird, dass das Leseverständnis sowie die Lernzielorientierung durch das Programm gefördert würde, sind diese positiv gerichtet formuliert. Die Vermeidungs-Leistungszielorientierung und die Präferenz zur Arbeitsvermeidung sind negativ gerichtet formuliert, da hier angenommen wird, dass das Programm diese reduziert. Die Hypothese der Leistungszielorientierung ist wiederum ungerichtet formuliert, da diese einerseits positive (kurzfristige) Auswirkungen auf den Lernerfolg haben kann, andererseits ein Abbau auch nicht lernhinderlich wäre; für weitergehende Informationen zu den theoretischen Fundierungen siehe Kapitel 4.3.

Tabelle 3 Fragestellungen & Hypothesen der quantitativen Analyse

Summative Evaluation		Formative Evaluation
Fragestellung	Hypothesen	Fragestellung
<p><i>2.1 Sind Verbesserungen in dem Leseverständnis der Kinder über die vier Jahre beobachtbar?</i></p>	<p>2.1.1 In der zweiten Klasse haben die Kinder ein höheres Leseverständnis als in der ersten Klasse.</p> <p>2.1.2 In der dritten Klasse haben die Kinder ein höheres Leseverständnis als in der zweiten Klasse.</p> <p>2.1.3 In der vierten Klasse haben die Kinder ein höheres Leseverständnis als in der dritten Klasse.</p>	<p><i>1.1 Welche Verläufe des Leseverständnisses lassen sich über die vier Jahre beobachten?</i></p>
<p><i>2.2 Gibt es positive Veränderungen in der motivationalen Orientierung der Kinder?</i></p>	<p>2.2.1 In der vierten Klasse haben die Kinder eine höhere Lernzielorientierung und niedrigere Vermeidungs-Leistungsziel- und Arbeitsvermeidungsorientierung als in der dritten Klasse.</p> <p>2.2.2 Die Annäherungs-Leistungszielorientierung der vierten Klasse unterscheidet sich von der dritten Klasse.</p>	

4.2 Stichprobe & Überblick Testungen

Zur Beantwortung der Fragestellungen werden zum Teil diejenigen Daten herangezogen, die ENB selbstständig im Verlauf der Förderzeit erhoben oder die ENB dem Autor dieser Studie zur Verfügung gestellt hat. Hierunter fallen einerseits die beschriebenen ELFE-Tests, die das Leseverständnis erheben und die einmal jährlich über die vier Schuljahre erhoben wurden sowie der SELMO-Test, der die motivationale Zielorientierung der Schülerinnen und Schüler in der dritten und vierten Klassenstufe testete. Andererseits wurden die Schuljahresnoten aus der zweiten bis vierten Jahrgangsstufe sowie diejenigen Kinder, die die Einzelförderung erhielten, der Evaluation zur Verfügung gestellt. Das bedeutet, dass im Rahmen dieser Evaluation keine Erhebung seitens des Autors dieser Studie mit den Kindern durchgeführt wurde.

Bei der Stichprobe, auf die sich die zur Verfügung stehenden Daten beziehen, handelt es sich demgemäß um die Schülerinnen und Schüler derjenigen Klasse, die durch ENB gefördert wurde. In Tabelle 4 wird die Anzahl der Kinder in der jeweiligen Klassenstufe dargestellt. Hierbei sind Fluktuationen über die Jahre erkennbar. Dies liegt einerseits daran, dass Schülerinnen und Schüler aus verschiedenen Gründen die Klasse verließen und

andererseits andere Schülerinnen und Schüler in die Klasse hinzukamen. Darüber hinaus wird auch das Geschlecht und die zuhause gesprochene Sprache in Tabelle 4 dargestellt. Gleichzeitig werden verschiedene Stichproben gegenübergestellt: Einerseits die gesamte Stichprobe, die aus allen Schülerinnen und Schülern besteht, die zu einem Zeitpunkt in der Klasse waren. Die bereinigte Stichprobe beinhaltet nur diejenigen Kindern, die den ELFE-Test zu allen Messzeitpunkten ausgefüllt hatten. Die Stichprobe SELLMO beinhaltet analog jene Kinder, die den SELLMO-Test zu allen Messzeitpunkten ausgefüllt hatten. Da bei der bereinigten Stichprobe sowie in der Stichprobe SELLMO jeweils nur Schülerinnen und Schüler mit vollständigen Datensätzen gewählt wurden, ist deren Anzahl über die Schuljahre hinweg konstant.

Tabelle 4 Schüleranzahl, Geschlecht und zuhause gesprochene Sprachen in den Stichproben

		Gesamte Stichprobe	Bereinigte Stichprobe	Stichprobe SELLMO
Anzahl	Gesamt	24	16	18
	1. Klasse	21	16	16
	2. Klasse	20	16	18
	3. Klasse	19	16	18
	4. Klasse	19	16	18
Geschlecht	Weiblich	18	14	16
	Männlich	6	2	2
Zuhause gesprochene Sprache	deutsch	2	2	2
	gemischt	16	12	13
	andere	6	2	3

Da abhängig der Stichprobe unterschiedliche Variablen betrachtet wurden, wird im Folgenden dargestellt, zu welchen Zeitpunkten welche Testungen durch das Programm durchgeführt wurden. In Tabelle 5 wird dargestellt, aus welcher Klassenstufe welche Daten entstammen. Die Noten beziehen sich auf die Jahresgesamtnoten; in der vierten Klassenstufe wurden darüber hinaus das Jahreszwischenzeugnis zur Verfügung gestellt, weshalb die Noten aus dem Jahreszwischen- und dem Jahresgesamtzeugnis vorliegen.

Tabelle 5 Erhebungszeitpunkte der quantitativen Längsschnittdaten

	ELFE	SELLMO	Noten
1. Klasse	x		
2. Klasse	x		x
3. Klasse	x	x	x
4. Klasse	x	x	2x

4. Klasse Noten: es stehen das Zwischen- und Abschlusszeugnis der 4. Klasse zur Verfügung

4.3 Erhebungsinstrumente

In Kapitel 4.3 wurde bereits beschrieben, welche Tests zu welchen Zeitpunkten durch das Programm erhoben wurden. Im Folgenden sollen die erhobenen Tests erläutert werden. Darüber hinaus werden knapp die theoretischen Modelle dargestellt, welche den Tests zugrunde liegen.

Das Programm verwendete einerseits den ELFE-Test (Lenhard et al., 2018) zur Erhebung des Leseverständnisses der Schülerinnen und Schülern. Die folgenden Informationen über den ELFE-Test entstammen ausschließlich aus Lenhard et al. (2018): Bei dem ELFE-Test handelt es sich nach Aussage dessen Autorinnen und Autoren um ein standardisiertes und normiertes Testverfahren von hoher Objektivität, Reliabilität und Validität. Der Anspruch des Tests ist die Erfassung des Leseverständnisses. In diesem Zuge werden drei Ebenen formuliert, die unterschiedliche basale und höhere Leseprozesse benötigen:

1. Erstens das **Wortverständnis**, welches das Dekodieren von Wörtern und die Synthese von Wortbestandteilen beinhaltet.
2. Die zweite Ebene bezieht sich auf das **Satzverständnis**, welches durch sinnennehmendes Lesen und syntaktische Fähigkeiten (Analyse von Satzbestandteilen) beschrieben wird.
3. Das **Textverständnis** beschreibt die letzte Ebene, die lokale und globale Kohärenzbildung umfasst. Hierbei wird ein Situationsmodell des Textes gebildet, welches Vorwissen und schlussfolgerndes Denken voraussetzt.

Das Satz- und Textverständnis stellen dabei höhere Anforderungen an die exekutiven Arbeitsgedächtnisfunktionen sowie das Vorwissen und das schlussfolgernde Denken als das Wortverständnis. Um diese drei Dimensionen zu messen, wurden verschiedene Items erstellt, welche den Kindern mit Zeitvorgaben im Rahmen der Testung vorgelegt werden. Die individuelle Leseleistung des Kindes ergibt sich aus den richtig gelösten Items pro Zeiteinheit. Nach Aussage der Autorinnen und Autoren des ELFE-Tests werden hiermit die Fähigkeiten gemessen, richtig zu lesen und die Aufgaben schnell zu bearbeiten. Das bedeutet, dass die Testergebnisse immer eine Interaktion zwischen Leseverständnis und Bearbeitungsgeschwindigkeit darstellen.

Nach Lenhard et al. (2018) ist eines der Hauptziele des ELFE-Tests die frühzeitige Diagnose von Lernschwierigkeiten beim Leseverstehen, damit Hinweise für daran anknüpfende Interventionsansätze generiert werden können. Durch den Fokus auf die drei beschriebenen Ebenen des Wort-, Satz- und Textverständnisses ist der Fokus des ELFE-

Tests sehr weit gefasst, mit dem Ziel, eine möglichst umfassende Diagnose hinsichtlich des Leseverständnisses zu bieten. Gleichzeitig eignet sich der Test zur Abbildung von Lernverläufen über große Zeiträume. Den Autorinnen und Autoren des ELFE-Tests nach ist der ELFE-Test heutzutage ein fester Bestandteil der Schulevaluationsforschung. Neben der Fragebogenkonstruktion wurde eine Normierungsstichprobe erhoben, die den pädagogischen Akteurinnen und Akteuren Referenzwerte bieten soll, um gewonnene Testergebnisse relational einordnen zu können. Die Normierungsstichprobe besteht aus 5073 Datensätze, welche in 9 verschiedenen Bundesländern der BRD erhoben wurde. Hierbei fungierten 36 Psychologiestudierende und Studierende des Lehramts als Testleiter, welche 71 Schulen und dabei insgesamt 351 Klassen akquirierten (darunter befindet sich allerdings nicht Bayern unter den Bundesländern). Die Erhebung erfolgte von März bis Dezember 2015. Aus den erhobenen Daten wurden zunächst unvollständige und unplausible Datensätze ausgeschlossen (weniger als 5 %) und anschließend wurden die Daten so selektiert, dass die Stichprobe nach den Kriterien der Schulart und des Geschlechts repräsentativ sind. Daten, die überrepräsentiert waren, wurden per Zufallsauswahl ausgeschlossen, um die anvisierte Quote zu erreichen. Aufgrund von datenschutzrechtlichen Bestimmungen erfolgte die Erhebung des Migrationshintergrundes nicht anhand der Eltern, sondern die Kinder wurden direkt im Fragebogen nach der zuhause gesprochenen Sprache befragt. Hierbei wurden drei Kategorien vorgegeben: Deutsch, gemischt (deutsch und mindestens eine andere Sprache) oder andere (Sprachen und nicht deutsch). In der Normstichprobe gaben 73,4 % der Kinder an, zuhause nur deutsch zu sprechen und 4,7 % der Kinder gaben an, zuhause gar kein deutsch zu sprechen. Nach Aussage von Lenhard et al. (2018) gilt die Stichprobe nach diesen Kennzahlen als repräsentativ hinsichtlich der Verteilung der zuhause gesprochenen Sprache und des daraus abgeleiteten Migrationshintergrundes. Zur genauen Berechnung der Normwerte und weiteren Details siehe Lenhard et al. (2018).

Die Zeitvorgaben bei der Durchführung der Tests sind abhängig der Untertests. Für jeden Untertests gibt es ein vorgegebenes Zeitintervall, wobei die Untertests jeweils mit spezifischen Instruktionen eingeleitet werden. Für die Bearbeitung der Untertests des Wort- und Satzverständnisses sind jeweils 3 Minuten vorgesehen und für das Textverständnis jeweils 7 Minuten. Insgesamt beträgt die reine Bearbeitungszeit der Testaufgaben somit 13 Minuten. Nach Aussage der Autorinnen und Autoren benötigt das Durchführen einer Gruppentestung mit allen nötigen Schritten zwischen 20 und 30 Minuten. Wichtig hierbei ist es, genau den Anweisungen zur Testdurchführung zu befolgen, die Lenhard et al. (2018) vorgeben, u. a. einen aufgeräumten Arbeitsplatz, keine inhaltlichen Antworten geben sowie möglichst Störungen während der Testung vermeiden. Denn die Ergebnisse haben nur dann Aussagekraft, wenn der Test unter standardisierten Bedingungen durchgeführt wird. Gleichzeitig beschreiben die Autorinnen und Autoren, dass die Kinder über hinreichende Sprach- und Buchstabenkenntnisse verfügen, da ansonsten die Leseleistung nicht zuverlässig erfasst werden kann. Die Ergebnisse der Tests (= Rohwerte) werden anschließend in T-Werte transformiert (Mittelwert = 50, Standardabweichung = 10). Allerdings werden nur T-Werte innerhalb des Intervalls 25-75 berichtet, da für Werte außerhalb des Intervalls keine ausreichend große Anzahl an Personen in der Normstichprobe vorliegt. Die drei Ebenen des Wort-, Satz- und Textverständnisses werden anschließend zu einem Gesamtwert überführt, der das durchschnittliche Leseverständnis abbilden soll. Dies liegt nach Aussage von Lenhard et al. (2018) daran, dass die Untertestergebnisse meistens recht homogen sind und deshalb durch einen Gesamtwert approximativ beschrieben werden

können. Allerdings können auch Unterschiede in den Untertestergebnissen auftreten, wobei die Autorinnen und Autoren zwei plausible Erklärungsstränge bieten:

- Einerseits könnte dies ein Hinweis darauf sein, dass einer oder mehrere Untertests nicht richtig durchgeführt wurde, aufgrund von falsch verstandenen Instruktionen oder Motivationsverlust.
- Andererseits könnte es auch auf tatsächliche Unterschiede in den Untertests zurückgeführt werden.

Aufgrund dessen haben die Autorinnen und Autoren Diskrepanzvergleiche in die automatische (computerbasierte) Auswertung der ELFE-Testergebnisse integriert: Diese überprüfen die Untertestergebnisse statistisch auf signifikante Unterschiede.

Andererseits setzte das Programm neben dem ELFE-Test auch den SELLMO-S-Test (Kurzform: SELLMO) ein (Spinath et al., 2012), welcher das Ziel hat, die motivationale Orientierung der Testpersonen zu erfassen. Die Autorinnen und Autoren des SELLMO-Tests wollen pädagogischen Fachkräften ein diagnostisches Instrument an die Hand geben, um insbesondere niedrige Ausprägungen in der Motivation erkennen und daran anknüpfend Empfehlungen und Maßnahmen ableiten zu können. In deren Argumentation lassen sich die Ursachen von nicht-fähigkeitsbedingten Leistungsmängeln in der Regel auf motivationale Faktoren zurückführen. Motivation wird in diesem Rahmen als Zielpräferenz definiert, insbesondere um Lern- und Leistungsunterschiede erklären zu können. Hierbei werden vier Arten der Zielorientierung unterschieden: *Lernzielorientierung* beschreibt die Präferenz einer Person, die eigenen Kompetenzen weiterentwickeln zu wollen. Laut Spinath et al. (2012) zeigen zahlreiche Studien zeigen, dass Lernziele langfristiges Engagement und gute Leistungen sichern. *Leistungszielorientierung* beschreibt jene Ziele, die mit dem Zeigen oder Nicht-Zeigen von Wissen und Können verbunden werden. Hierbei wird die Leistungszielorientierung unterschieden in die Annäherungs- und Vermeidungs-Leistungszielorientierung. Annäherungs-Leistungszielorientierung beschreibt die Präferenz, eigenes Wissen und Können anderen Personen zeigen zu wollen. Diese geht kurzfristig mit guten Leistungen einher, sichert aber nicht die andauernde, langfristige Beschäftigung mit Themen, die für andauernde Erfolge nötig ist. Vermeidungs-Leistungszielorientierung beschreibt wiederum, dass eigenes mangelndes Wissen und Können nicht vor anderen gezeigt werden will. Dies geht kurz- und langfristig mit schlechten Leistungen einher. Als vierte Zielkategorie beschreiben die Autorinnen und Autoren die Präferenz zur *Arbeitsvermeidung*. Hierbei besteht weder ein Lern- noch ein Leistungsanreiz, sondern die Person verfolgt lediglich das Ziel, möglichst wenig Arbeit zu investieren. Dies geht kurz- und langfristig mit geringem Engagement und schlechten Leistungen einher

Der SELLMO-Test verfügt ebenfalls über eine Normstichprobe, die in diesem Zusammenhang als Eichstichprobe bezeichnet wird. Diese bestand aus 3348 Schülerinnen und Schüler der dritten bis zehnten Klassenstufe. Die Erhebung fand im Schuljahr 2011/2012 statt. 291 Schülerinnen und Schüler stammen hierbei aus der dritten Klasse und 361 Schülerinnen und Schüler aus der vierten Klasse und somit beinhaltet die Normstichprobe 652 Kinder aus der Grundschule. Allerdings fand keine Erhebung der Normstichprobe in Bayern statt. Die Testdurchführung beschreiben die Autorinnen und Autoren analog zum ELFE-Test, wobei hier keine Zeitvorgaben vorgegeben werden. Voraussetzung für die Testbearbeitung ist ein geeigneter Entwicklungs- und Leistungsstand der Kinder sowie ein

ausreichendes Sprachverständnis. Die Testergebnisse werden ebenfalls als T-Werte berichtet.

4.4 Methoden

Im Folgenden werden nun diejenigen inferenzstatistischen Methoden beschrieben, mit welchen die Fragestellungen der summarischen Evaluation beantwortet werden sollen. Hierzu wurde einerseits eine einfaktorielle Varianzanalyse mit Messwiederholungen (**rmANOVA**) sowie ein t-Test für abhängige Stichproben berechnet. Die Datenanalyse und die erstellten Grafiken wurden mit dem Statistikprogramm R durchgeführt (R Core Team, 2024). Das Signifikanzniveau aller inferenzstatistischer Tests wurde auf $\alpha < 0.5$ gesetzt (Bortz & Schuster, 2010). Falls nun ein p-Wert kleiner als 5 % liegt, wird somit von einem statistisch signifikanten Ergebnis gesprochen. Die rmANOVA untersucht die Hypothesen in Bezug auf die ELFE-Testergebnisse über alle vier Schuljahre. Zunächst wurde überprüft, ob die Daten bestimmte Voraussetzungen erfüllen (siehe Bortz & Schuster, 2010): Dabei wird ein Shapiro-Wilk-Test durchgeführt, um die Normalverteilung der abhängigen Variable zu jedem Messzeitpunkt zu überprüfen. Falls der Shapiro-Wilk-Test signifikant wird ($p < 0.5$), kann man davon ausgehen, dass keine Normalverteilung der Daten gegeben ist. Darüber hinaus wurde die Sphärizität und das Vorhandensein von Ausreißern überprüft (Bortz & Schuster, 2010). Ausreißer werden definiert ab 1,5-fachen Interquartilsabstand unterhalb des ersten Quartils oder oberhalb des dritten Quartils (Bortz & Schuster, 2010). Besteht ein 3-facher Abstand, wird von einem extremen Ausreißer gesprochen. Zur Überprüfung der Sphärizität wurde der Mauchly-Test eingesetzt. Da zur Durchführung der rmANOVA das rstatix-Paket in R verwendet wurde, wurde kein separater Mauchly-Test berechnet, da dieser bereits im Paket enthalten ist und bei Verletzung der Voraussetzung der Sphärizität automatisch angepasst wird (Kassambara, 2023). Nach abschließender Überprüfung der Voraussetzungen wurde für den Signifikanztest eine F-Statistik über alle Messzeitpunkte berechnet. Handelt es sich bei der F-Statistik um einen Wert, der unter Angabe der Freiheitsgrade in der entsprechenden F-Verteilung bei Gültigkeit der Nullhypothese einen kleineren p-Wert als 0.5 bekommt, so spricht man von einem signifikanten Unterschied. Das bedeutet, dass sich mindestens zwei der Messzeitpunkte statistisch signifikant voneinander unterscheiden. Da hierdurch allerdings nicht ersichtlich wird, welche Messzeitpunkte sich statistisch signifikant voneinander unterscheiden, werden anschließend Post-hoc-Tests berechnet (Bortz & Schuster, 2010). Hierzu werden alle Messzeitpunkte miteinander auf signifikante Unterschiede mittels t-Tests verglichen. Aufgrund der Alphafehlerkumulierung werden die p-Werte mit der Bonferroni-Korrektur korrigiert, weshalb ein adjustierter p-Wert berechnet wird, der zur Interpretation der statistischen Signifikanz herangezogen wird.

Zur Berechnung der **t-Tests** für abhängige Stichproben wurden ebenfalls zunächst überprüft, ob der Datensatz die jeweiligen Voraussetzungen erfüllt. Hierbei wurde insbesondere die Voraussetzung der Normalverteilung jedes Messzeitpunktes mittels eines Shapiro-Wilk-Tests überprüft (Bortz & Schuster, 2010). Hierbei wird die Differenz beider Mittelwerte berechnet und die sich daraus ergebende Verteilung mittels des Shapiro-Wilk-Tests auf Normalverteilung hin überprüft. Falls die Voraussetzung erfüllt wird, wird eine t-Statistik berechnet, die ähnlich zum F-Test mit der t-Verteilung verglichen wird. Sollte ein p-Wert unter 0.5 herauskommen, kann man von einem statistisch signifikanten Ergebnis sprechen. Sollte die Voraussetzung der Normalverteilung nicht erfüllt sein, so wird auf das nonparametrische Verfahren des Wilcoxon-Vorzeichen-Rang-Tests zurückgegriffen, welcher

nicht der Voraussetzung der Normalverteilung bedarf. Hierbei werden für jede Person aus den Differenzen zwischen den Messzeitpunkten Ränge bestimmt und daraus mittels der jeweiligen Rangsummen die Teststatistik W berechnet, mit welcher wiederum ein z -Wert berechnet wird, der auf statistische Signifikanz überprüft wird. Für alle Tests werden auch Effektgrößen bestimmt. Dies geschieht nach Cohen (1988), indem die Mittelwertsdifferenz an der gepoolten Standardabweichung relativiert wird. Die Effektgrößen der Ergebnisse der inferenzstatistischen Analysen werden im Ergebnisteil ebenfalls nach Cohen (1988) berichtet: Dieser beschreibt Effektgrößen von > 0.1 als klein, > 0.3 als mittel und > 0.5 als groß. Sollte der Wilcoxon-Vorzeichen-Rang-Test eingesetzt worden sein, so werden die Effektgrößen folgendermaßen berechnet: Eine Effektgröße r wird als z -Statistik berechnet, indem der Z -Wert durch die Wurzel der Stichprobengröße geteilt wird (Pallant, 2020). Die daraus resultierende Effektgröße wird nach Cohen (1988) eingeschätzt.

4.5 Deskriptive Analysen

Im Folgenden werden die deskriptiven Analysen dargestellt. Dies bezieht sich allein auf die in Kapitel 4.3 dargestellte Schülerstichprobe. Deskriptive Analysen enthalten keine mittels inferenzstatistischer Methoden gewonnener Ergebnisse. Es werden lediglich Häufigkeitsverteilungen dargestellt. Diese sollen einerseits einen Überblick über die Ergebnisse der verwendeten quantitativen Instrumente bieten und andererseits den Programmbeteiligten von ENB Ansatzpunkte für eigene Interpretationen geben.

Da im Rahmen einiger statistischer Tests der Datensatz um fehlende Werte bereinigt werden musste, werden folglich die deskriptiven Werte der Gesamtstichprobe mit denen der bereinigten Stichprobe gegenübergestellt. Die bereinigte Stichprobe umfasst diejenigen Kinder, die in den ELFE-Testergebnissen und auf den Noten kein Missing besitzen, wodurch für jedes Kind dieser Stichprobe auf jedem Messzeitpunkt ein Wert vorhanden ist. Zu diesem Vergleich kommt eine weitere, reduzierte Stichprobe hinzu, die im Rahmen der Analyse der Testergebnisse des SELLMO-Tests verwendet wurde. Diese besteht aus denjenigen Schülerinnen und Schülern, die in der dritten und vierten Klasse Werte im SELLMO-Test aufweisen, sodass für jede Schülerin und jeden Schüler in der dritten und vierten Klasse in dem SELLMO-Test ein Wert vorhanden ist. Zusammenfassend werden in den deskriptiven Statistiken demzufolge drei Stichproben gegenübergestellt: Die Gesamtstichprobe, eine bereinigte Stichprobe über alle Messzeitpunkte und eine Stichprobe für den SELLMO-Test. Erläuterung zuhause gesprochene Sprache: Wie in Kapitel 4.3 beschrieben, wurde bei der Frage nach den zuhause gesprochenen Sprachen drei Kategorien gebildet: „Deutsch“ bedeutet, dass zuhause ausschließlich deutsch gesprochen wird. „Gemischt“ bedeutet, dass zuhause sowohl deutsch als auch eine andere Sprache gesprochen wird und „andere“ bedeutet, dass zuhause kein deutsch gesprochen wird. In den deskriptiven Statistiken werden die Daten nicht auf dem höchsten Detaillierungsgrad dargestellt. So wird aufgrund der Übersichtlichkeit lediglich die Anzahl der Schülerinnen und Schüler für jede Klassenstufe dargestellt. Aufgrund derselben Argumentation wird auf die deskriptive Darstellung möglicher Merkmalskombination wie Geschlecht und zuhause gesprochener Sprachen – auch in den inferenzstatistischen Analysen – verzichtet.

4.5.1 Soziodemographische und schul- und programmbezogene Daten

Im Rahmen des ELFE-Tests wurden Informationen zu dem Schülergeschlecht und der zuhause gesprochenen Sprache erhoben (siehe Kapitel 4.3). Darüber hinaus wurden dem Autor dieser Arbeit durch die Programmleitung von ENB Informationen über die Häufigkeit bzw. Intensität der Einzelförderung einzelner Kinder mitgeteilt, die gleichfalls dargestellt werden. Des Weiteren liegen Informationen vor, ob ein Schüler oder eine Schülerin die erste Klasse wiederholte. Seitens der Programmleitung wurde mitgeteilt, dass die Kinder mit dem anonymisierten Code 110 sowie 117 die erste Klassenstufe wiederholten. In Tabelle 6 werden diejenigen Kinder aufgelistet, die durch das Programm im Rahmen der Einzelförderung gefördert wurden. Die Angaben entstammen von der Programmleitung, wobei die Kategorien, die sich auf die Häufigkeiten beziehen, nicht genauer spezifiziert wurden und lediglich eine Rangfolge darstellen sollen. „In der zweiten Klasse“ bezieht sich auf jene Kinder, die lediglich in der zweiten Klasse und darüber hinaus nicht mehr oder lediglich stark reduziert an der Einzelförderung teilnahmen. Alle anderen Kategorien beziehen sich auf alle Förderjahre.

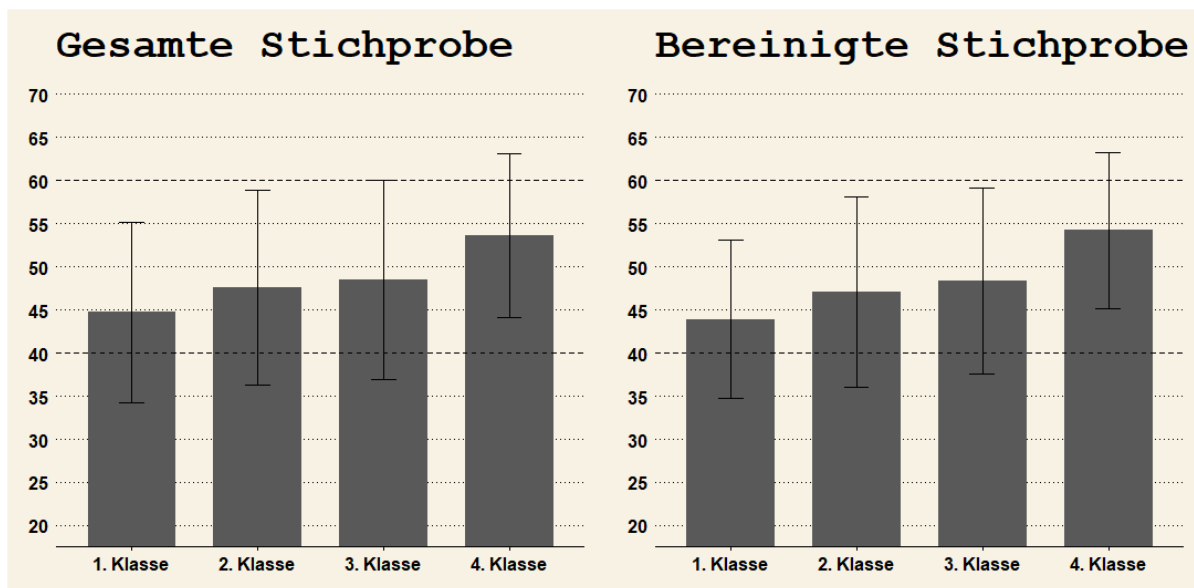
Tabelle 6 Einzelgeförderte Kinder

	Gelegentlich	Regelmäßig	Wöchentlich	In der zweiten Klasse
	103	107	106	104
Einzelgeförderte	105	109	110	123
Kinder	108		113	
	124		116	

4.5.2 Leistungsrelevante Daten

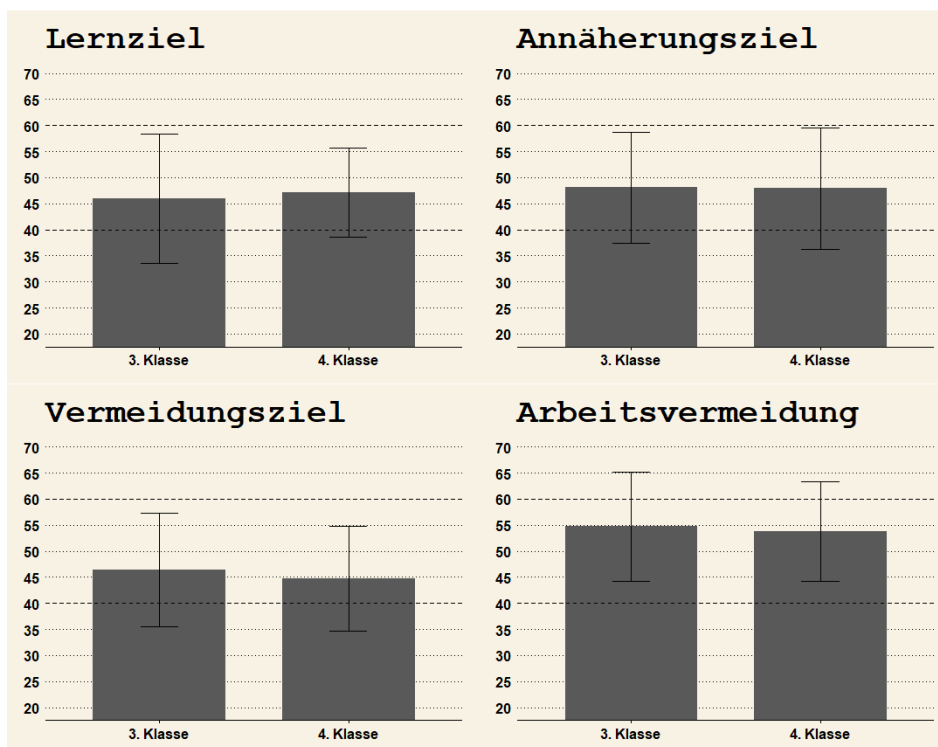
Im Folgenden werden die durch die eingesetzten quantitativen Instrumente erhobenen Testergebnisse dargestellt. Hierbei werden zunächst die Ergebnisse der **ELFE**-Tests der gesamten Stichprobe und der bereinigten Stichprobe angegeben. Aus Übersichtlichkeitsgründen werden die Ergebnisse der Stichprobe SELLMO nicht dargestellt, insbesondere, da sie sich nicht stark von denen der bereinigten Stichprobe unterscheiden. Die Testergebnisse der ELFE-Test werden in dieser Arbeit lediglich im Rahmen der Gesamtwerte berichtet und nicht der einzelnen Untertestergebnissen. Eine ausführliche diesbezügliche Argumentation findet sich in Kapitel 4.6.1. In Abbildung 4 werden die Gesamtwerte des mit dem ELFE-Test erhobenen Leseverständnisses über die vier Erhebungszeitpunkte graphisch dargestellt. Hierbei werden die T-Werte dargestellt (siehe Kapitel 4.5). Die gestrichelten Linien markieren den von ELFE definierten Normalbereich, der innerhalb einer Standardabweichung um den Mittelwert der Normstichprobe liegt.

Abbildung 4 Mittelwerte und Standardabweichung der Gesamtwerte der ELFE-Testergebnisse



In Abbildung 5 werden die **SELLMO**-Testergebnisse visuell dargestellt. Da hier keine Gesamtskala existiert, werden die jeweiligen Skalen einzeln aufgeführt. Die gestrichelten Linien geben den durch SELLMO vorgegebenen Normalbereich an. Es werden wiederum die T-Werte der Testergebnisse dargestellt (siehe Kapitel 4.3).

Abbildung 5 Mittelwerte und Standardabweichung der SELLMO-Testergebnisse



Die **Klassendurchschnittsnoten** der Schülerinnen und Schüler werden als Gesamtdurchschnittsnote, die im Jahreszeugnis aufgeführt wurde, dargestellt. Aufgrund des starken Fokus des Programms auf sprachliche Inhalte werden darüber hinaus die durchschnittlichen Klassennoten im Unterrichtsfach Deutsch der Gesamtdurchschnittsnote gegenübergestellt. Da in der ersten Klassenstufe keine Noten vergeben wurden, beginnt die Datenreihe ab dem zweiten Schuljahr. Des Weiteren stehen für diese Auswertung die Noten aus dem Zwischenzeugnis der vierten Klassenstufe zur Verfügung, die einen differenzierteren Blick in das vierte Schuljahr erlauben. Diese werden ebenfalls in die deskriptiven Analysen aufgenommen. In Abbildung 6 werden die Gesamtdurchschnittsnoten („Notendurchschnitt“) sowie die Deutschnote der Klasse über die Schuljahre hinweg aufgeführt. Dies betrifft die gesamte Stichprobe. In Abbildung 7 wird darüber hinaus zum Vergleich mit der gesamten Stichprobe die Notenwerte der bereinigten Stichprobe aufgeführt.

Abbildung 6 Mittelwert und Standardabweichung der Noten der gesamten Stichprobe

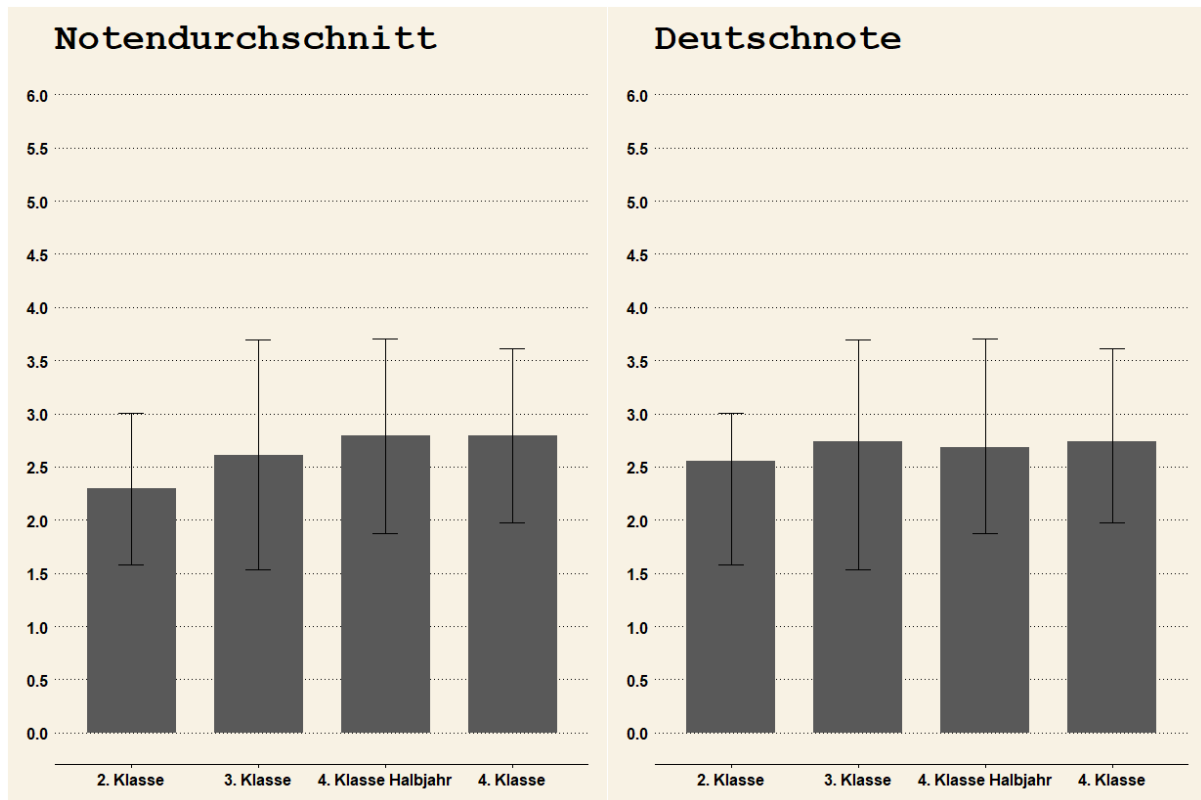
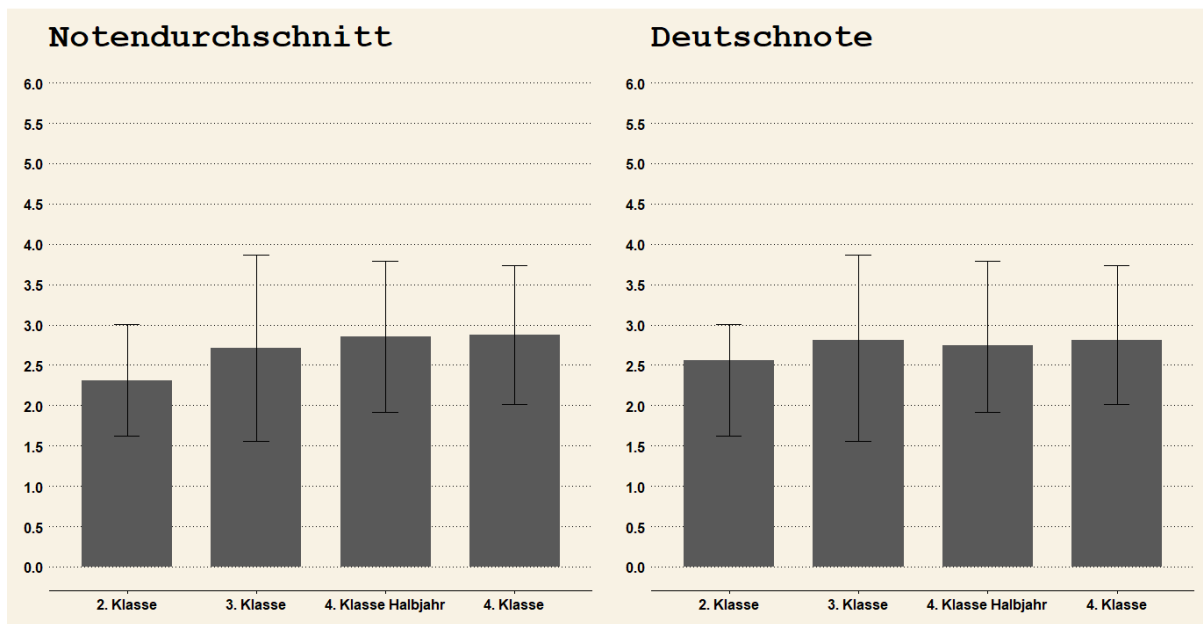


Abbildung 7 Mittelwerte und Standardabweichungen der Noten der bereinigten Stichprobe



Anhand der Abschlussnote des vierten Schuljahres lässt sich die Durchschnittsnote der Schülerinnen und Schüler berechnen. Diese Note bestimmt die Möglichkeit des **Übertritts** auf eine weiterführende Schule. Da nach Aussage der Programmleitung jedes Kind, dass eine entsprechende Note für das Gymnasium erlangte, auch eine Übertrittsempfehlung seitens der Lehrkraft ausgesprochen bekam, werden im Folgenden lediglich die durchschnittlichen Abschlussnoten betrachtet und davon ausgehend wird die Möglichkeit eruiert, einen Übertritt auf die Realschule oder auf das Gymnasium realisieren zu können. Dies wurde folgendermaßen bestimmt: Jene Schülerinnen und Schüler, die einen besseren Notendurchschnitt am Ende der vierten Klasse von 2,33 erwarben, haben die Möglichkeit, ein Gymnasium zu besuchen. Entgegengesetzt ist denjenigen Kindern, die einen schlechteren Notendurchschnitt als 2,66 erwarben, der Zutritt zur Realschule und dem Gymnasium zunächst mittels Durchschnittsnote und Übergangsempfehlung nicht möglich. In Tabelle 7 werden die entsprechenden Übertrittsquoten sowie weiterführende Merkmale im Zusammenhang dargestellt. So wird des Weiteren die Übertrittsnote nach dem Geschlecht, der zuhause gesprochenen Sprache sowie der Teilnahme an der Einzelförderung aufgegliedert und tabellarisch dargestellt. Die Daten werden sowohl für die gesamte Stichprobe als auch die bereinigte Stichprobe dargestellt.

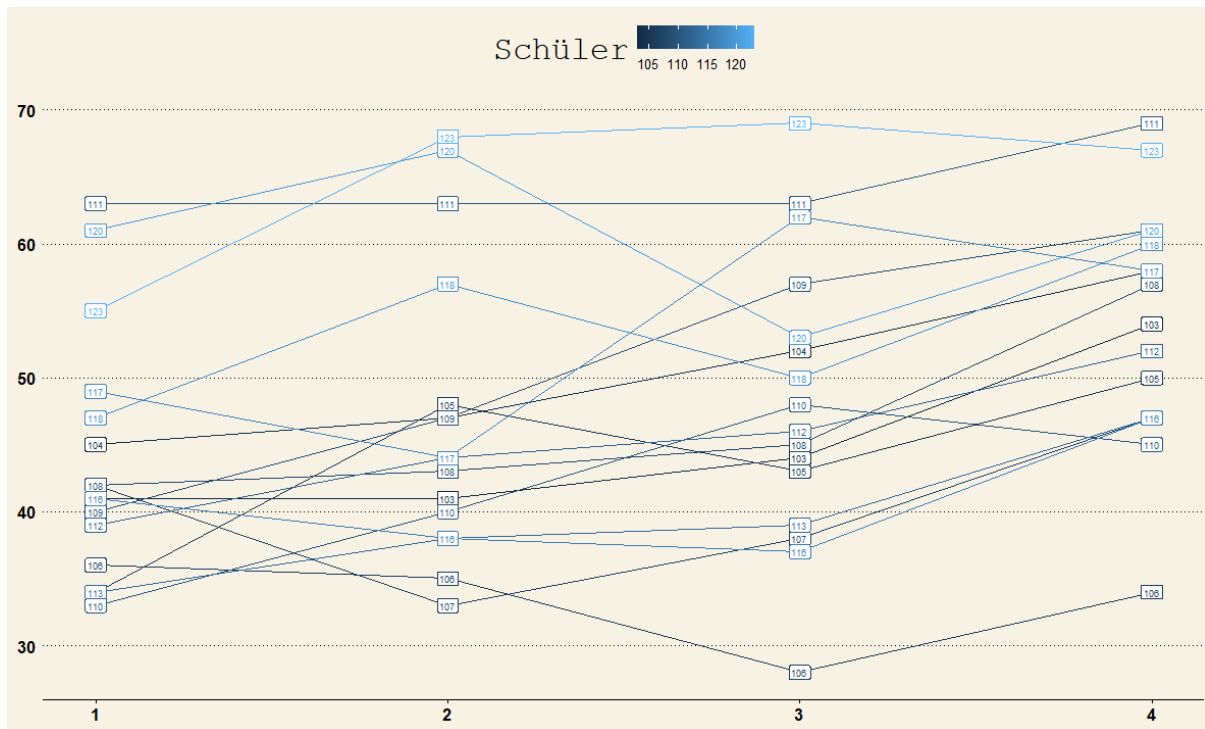
Tabelle 7 Deskriptive Statistiken der Übertrittsnoten auf weiterführende Schulen

		Gesamte Stichprobe		Bereinigte Stichprobe	
		Gymnasium	Mittelschule	Gymnasium	Mittelschule
Anzahl	Gesamt	9	9	7	9
Geschlecht	Weiblich	7	9	5	9
	Männlich	2	0	2	0
Zuhause gesprochene Sprachen	deutsch	0	2	0	2
	gemischt	7	6	6	6
	andere	2	1	1	1
Einzelgeförderte Kinder	Gelegentlich	0	3	0	3
	Regelmäßig	0	2	0	2
	Wöchentlich	0	4	0	4
	In der zweiten Klasse	2	0	2	0

4.5.3 Individuelle Lernverläufe

In den bisherigen Analysen wurden ausschließlich die Klassenmittelwerte und -standardabweichungen betrachtet und vom einzelnen Kind abstrahiert. Um einen Einblick in die individuellen Lernverläufe zu bekommen, sind in Abbildung 8 die Lernverläufe dargestellt, die sich aus den ELFE-Gesamtwerten über die einzelnen Erhebungszeitpunkte ergaben. Dies soll einen differenzierten Blick auf die Veränderungen des Leseverständnisses über die Zeit geben. Des Weiteren soll auf die Darstellung der Veränderungen der Noten über die Zeit verzichtet werden, da aufgrund der hohen Häufung gleicher Notenwerte die Interpretation äußerst erschwert wird. Sollte dennoch Interesse bestehen, so ist die entsprechende Abbildung im Anhang (Individuelle Lernverläufe – Jahresdurchschnittsnote) aufgeführt. Weitere Merkmale wie die SELMO-Testergebnisse werden ebenfalls nicht dargestellt, allerdings auch nicht im Anhang aufgeführt. Dies hat den Grund, dass dies in einer großen Anzahl an Abbildungen resultieren würde, die der Übersichtlichkeit abträglich wären. Gleichzeitig soll der Schwerpunkt auf längerfristige Veränderungen gelegt werden und nicht allein auf die Veränderungen über zwei Messzeitpunkte.

Abbildung 8 T-Werte der einzelnen Kinder über die vier Messzeitpunkte



4.6 Inferenzstatistische Analysen

4.6.1 Diskrepanzvergleiche zwischen den ELFE-Untertests

Wie bereits beschrieben besteht der ELFE-Test aus drei Untertests (Wort-, Satz- und Textverständnis). Da die Ergebnisse der Untertests laut Lenhard et al. (2018) in der Regel nicht stark voneinander abweichen, lässt sich der Gesamtmittelwert über die drei Untertests hinweg heranziehen. Unterschiede der Untertests: Falls die Untertestergebnisse voneinander abweichen, nennen die Autorinnen und Autoren zwei Interpretationsansätze (Lenhard et al., 2018): Einerseits können Unterschiede ein Hinweis darauf sein, dass einer oder mehrere der Untertests nicht richtig durchgeführt worden ist (z. B. aufgrund von Motivationsverlust während der Testung oder falsch verstandene Instruktionen). Andererseits können aber auch tatsächliche Unterschiede zwischen den Untertests vorliegen, die die Unterschiede in den Ergebnissen bewirken. Aufgrund dessen wurden in die standardisierte Auswertung von ELFE Signifikanztests eingebaut, die die Diskrepanzvergleiche zwischen den Untertests statistisch auf Signifikanz überprüfen. Das Signifikanzniveau ist auf 5 % festgesetzt. Ein signifikantes Ergebnis würde bedeuten, dass der Unterschied zwischen den T-Werten größer ist, als man wegen des Messfehlers erwarten würde (Lenhard et al., 2018). Die Interpretation der signifikanten Diskrepanzvergleiche erfolgt vor dem Hintergrund der Einschätzung des Einflusses von Störvariablen auf die Messung. Ist davon auszugehen, dass ein großer Messfehler vorliegt – insbesondere aufgrund des nicht-standardisierten Vorgehens während der Testung und/oder aufgrund von getätigten Beobachtungen während der Testung, welche auf mögliche Einflussursachen hinweisen könnten – so sind die Signifikanzen der Diskrepanzvergleiche mit Zurückhaltung zu interpretieren (vgl. Lenhard et al., 2018). Wenn nun ein signifikanter Unterschied vorliegt, dann bietet das Gesamtergebnis, das den Durchschnittswert über alle Messungen hinweg angibt, kein reliables und/oder valides Gesamtmaß des Leseverständnisses (vgl. Lenhard et al., 2018).

Zur Interpretation von signifikanten Diskrepanzvergleichen schlagen Lenhard et al. (2018) vor, als nächsten Schritt zu analysieren, ob es sich im Vergleich mit der Normstichprobe um eine auffällige Grundrate handelt. Dabei empfehlen diese lediglich Grundraten zu betrachten, die unter 5 % liegen. Dies bedeutet, dass die vorgefundene Diskrepanz oder eine größere in weniger als 5 % aller Fälle der Normstichprobe vorkam (Lenhard et al., 2018). In diesem Fall würde es sich nach Aussage der Autorinnen und Autoren um eine diagnostisch bedeutsame Diskrepanz handeln, die auf erwartungswidrige Unterschiede zwischen den Untertests hindeuten würde. Folglich wäre dies ein Hinweis darauf, dass der Gesamtwert für das Leseverständnis nicht reliabel und/oder valide ist und dass die Untertests für sich betrachtet und auf ihre Aussagekraft hin überprüft werden sollten. Aufgrund dessen werden im Rahmen der Analyse der Gesamtwerte des Leseverständnisses zunächst die Diskrepanzvergleiche auf Signifikanz und anschließend auf deren Grundrate hin überprüft. Hierbei sollen jene Ergebnisse identifiziert werden, die sowohl signifikant sind als auch eine Grundrate von unter 5 % aufweisen. Abbildung 9 stellt die Häufigkeitsverteilungen dieser Ergebnisse dar. Im Folgenden sollen nun diese Ergebnisse interpretiert werden, um für oder gegen die Nutzung der Gesamtwerte in den statistischen Analysen zu argumentieren. In diesem Kontext werden ausschließlich die Häufigkeitsverteilungen betrachtet und nicht die Richtung der Diskrepanzen, da diese bereits visuell inspiziert worden sind und dem Autor dieser Studie in diesem Zuge keine Auffälligkeiten auffielen.

1. Klasse:

In der ersten Klassenstufe liegen die meisten signifikanten Ergebnisse mit einer Grundrate unter 5 % im Bereich der Wort-Satz- und Wort-Text-Diskrepanzen vor. Dies könnte dahingehend interpretiert werden, dass am Ende der ersten Klasse zwar schon Kenntnisse im Wortverständnis, aber noch fehlende Kenntnisse im Satz- und Textverständnis vorkommen. Gleichzeitig lässt sich vorstellen, dass die Schülerinnen und Schüler bei der Durchführung der Tests eventuell die Aufgabenstellung und/oder die Instruktionen zu den Satz- und Textverständnisaufgaben nicht hinreichend verstanden. Dies drückt sich darin aus, dass es zwischen den Satz-Text-Diskrepanzen vergleichsweise wenig signifikante Ergebnisse mit einer Grundrate von unter 5 % befinden.

2. Klasse:

In der zweiten Klassenstufe dreht sich die Verteilung visuell um eine vertikale Achse. Hierbei scheinen sich die Wort-Satz-Diskrepanzen reduziert zu haben, was darauf hindeuten könnte, dass die Aufgabenstellungen des Wort- und Satzverständnisses entweder aufgrund von besseren Fähigkeiten oder aufgrund eines besseren Verständnisses der Instruktionen oder Aufgabenstellung leistungsadäquater und einheitlicher bearbeitet wurden. Allerdings scheinen die Wort- und Satzverständnisaufgaben noch mit den Textverständnisaufgaben zu diskrepieren, was auf mangelnde Fähigkeiten zurückzuführen sein könnte.

3. Klasse:

In der dritten Klassenstufe lässt sich eine Gleichverteilung beobachten, was darauf hindeutet, dass alle Aufgaben ungefähr gleich gut verstanden wurden und es dennoch bei vereinzelt Kindern zu erwartungswidrigen Unterschieden in den Untertests gekommen ist. Im Vergleich zu den anderen Erhebungszeitpunkten liegen hier allerdings die geringsten Häufigkeiten in Kombination der Ausprägungen vor.

4. Klasse:

In der vierten Klassenstufe lässt sich eine Verdopplung der Anzahl der Wort-Text-Diskrepanzen beobachten. Dies ließe sich eventuell dadurch erklären, dass den Kindern das Bearbeiten der Wortverständnisaufgaben teilweise zu einfach oder zu simpel vorkam und deshalb Motivations- oder Interessensmängel die Unterschiede hervorrufen könnten. Da die Wort-Satz- und Satz-Text-Diskrepanzen auf einem gleichen Niveau wie in der 3. Klasse liegen, scheint es zu keinen großen Veränderungen gekommen zu sein.

Zusammenfassend lässt sich sagen, dass für die jahrgangsspezifischen Häufigkeitsverteilungen plausible Alternativerklärungen angeben lassen, die eher weg von dem Vorliegen tatsächlicher Unterschiede in den Untertests weisen. Darüber hinaus soll an dieser Stelle angemerkt werden, dass die Erhebungen der ELFE-Tests teilweise nicht nach standardisierten Bedingungen erfolgte, was durch den Versuchsleiter demzufolge begründet wurde, dass die Kinder aufgrund von Sprach- und Verstehensschwierigkeiten einige Items oder Aufgabenstellungen nicht sofort verstanden. Dies wäre ein weiterer Indikator dafür, dass die Diskrepanzen eher auf untersuchungsbedingte Messfehler zurückzuführen sind als auf tatsächlich vorliegende Unterschiede.

Abbildung 9 Signifikante Diskrepanzvergleiche



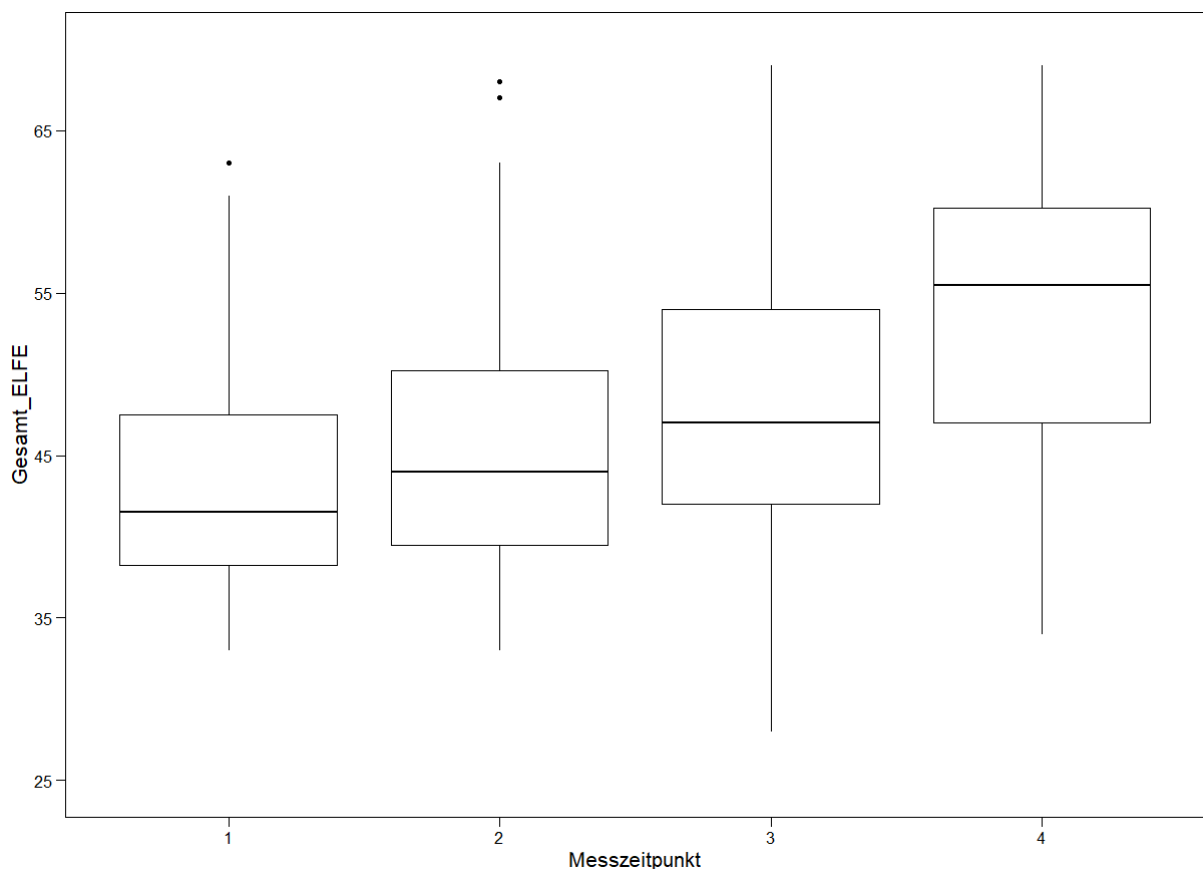
Signifikanzniveau von 5 %; Darstellung jener Ergebnisse, die eine Grundrate von < 5% aufweisen

4.6.2 Schülerleistungen über die Zeit

Im Folgenden sollen nun die Hypothese der Fragestellung 2.1 (siehe Kapitel 3.3) inferenzstatistisch anhand der gegebenen empirischen Daten überprüft werden. Hierzu wird auf die Methode der rmANOVA – wie in Kapitel 4.4 beschrieben – zurückgegriffen. Dabei wird auf die bereinigte Stichprobe zurückgegriffen, da die rmANOVA vollständige Datensätze über alle Messzeitpunkte bedarf. Die rmANOVA analysiert demzufolge, ob sich mindestens zwei Klassenmittelwerte über alle Erhebungszeitpunkte statistisch signifikant voneinander unterscheiden.

Um allerdings zu überprüfen, ob sich der Datensatz eignet, eine rmANOVA zu berechnen, müssen zunächst die Voraussetzungen überprüft werden. Diese werden in Kapitel 4.6 beschrieben und im Folgenden kurz dargestellt. Um zu identifizieren, ob sich **Ausreißer** in den Daten befinden, werden diese in Abbildung 10 als Boxplots dargestellt. Bei jenen Punkten, die einzeln abgebildet sind, handelt es sich um Datenpunkte, die über den 1,5-fachen Interquartilsabstand des dritten Quantils entfernt liegen. Ab einer Distanz von mehr als einem 3-fachen Interquartilsabstand spricht man von extremen Ausreißern. Bei den in Abbildung 10 dargestellten Punkten handelt es sich demzufolge nicht um extreme Ausreißer nach dieser Definition. Sie beschreiben die Schülerinnen und Schüler: 110, 120 sowie 123. Um sicherzugehen, dass die Ausreißer die Analyse nicht zu stark verzerren, werden zwei Analysen berechnet, jeweils mit und ohne diese Ausreißer.

Abbildung 10 Box-Plots zur Identifikation von Ausreißern der ELFE-Testergebnisse



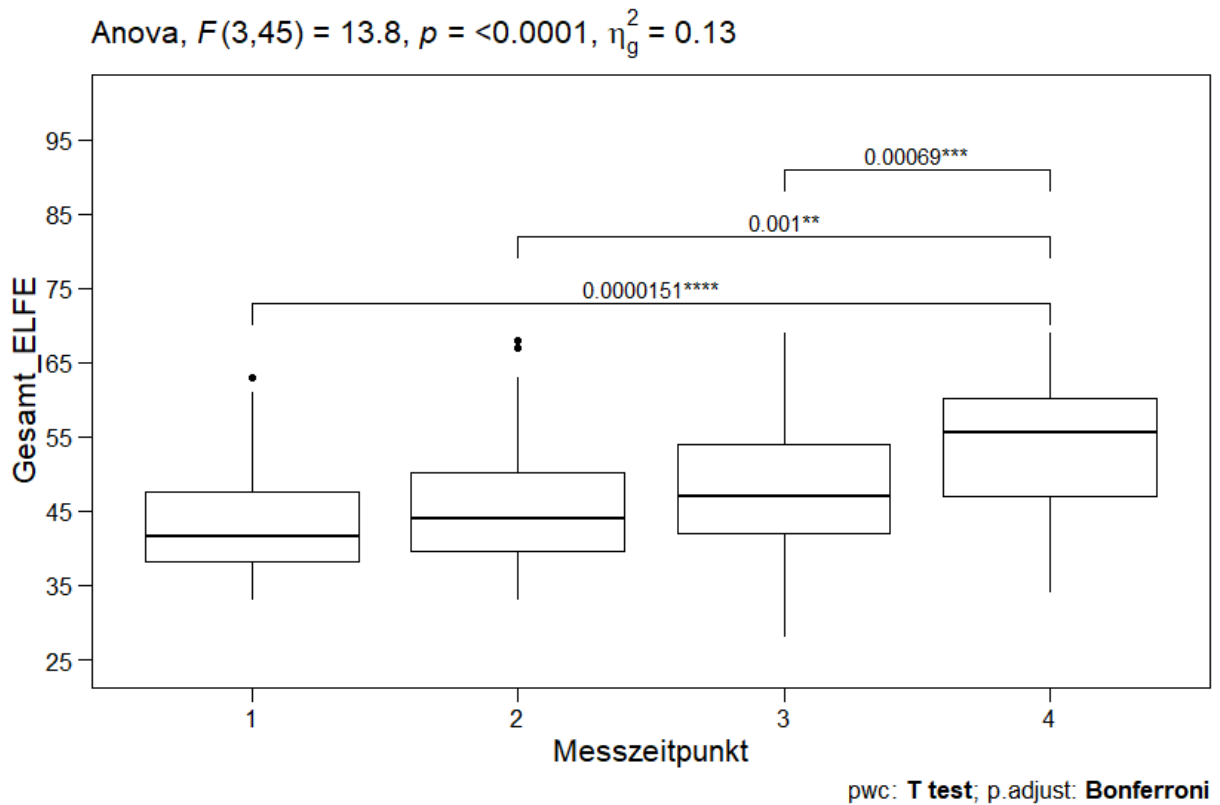
Die **Normalverteilung** der Variable zu jedem einzelnen Messzeitpunkt wurde mittels des Shapiro-Wilk-Tests berechnet. Wie in Tabelle 8 ersichtlich, wird keiner der Shapiro-Wilk-Test signifikant, wonach keine Verletzung der Normalverteilungsannahme angenommen wird.

Tabelle 8 Ergebnisse des Shapiro-Wilk-Tests der ELFE-Testergebnisse

Messzeitpunkt	Statistik	p
1	0.897	0.0742
2	0.888	0.0533
3	0.984	0.989
4	0.966	0.772

Die Berechnung der **rmANOVA** ergab, dass der Gesamtwert des ELFE-Tests über alle Zeitpunkte hinweg statistisch signifikant war, $F(3, 45) = 13.798$, $p = 0.00000163$, generalisiertes Eta-Quadrat = 0.129. Gerichtete Post-hoc Analysen mit einer Bonferroni-Korrektur zeigten auf, dass die Unterschiede zwischen dem ersten und dem vierten Zeitpunkt, $t(15) = -6.90$, $p_{adj} = 0.0000151$, $d =$), dem zweiten und dem vierten Zeitpunkt, $t(15) = -4.51$, $p_{adj} = 0.001$, $d =$) sowie dem dritten und dem vierten Zeitpunkt ($p_{adj} = 0.00069$) statistisch signifikant voneinander unterschiedlich waren (siehe Abbildung 11). Hierbei war der Gesamtwert jeweils beim vierten Zeitpunkt größer ausgeprägt als bei den anderen Zeitpunkten.

Abbildung 11 Box-Plots mit signifikanten Post-hoc-Tests der ELFE-Testergebnisse



Gerichtete Post-Hoc Tests: siehe Hypothesen

In Tabelle 9 werden die vollständigen Ergebnisse der Post-hoc Analysen dargestellt. Nach Cohen (1988) handelt es sich ab $d > 0.5$ um einen großen Effekt. Bei allen Post-hoc-Tests, bis auf den Vergleich 2-3 handelt es sich mindestens nach Cohen (1988) um einen großen Effekt. Bei dem Vergleich 2-3 handelt es sich wiederum um einen kleinen Effekt, da er kleiner als 0.1 ist.

Tabelle 9 Post-hoc-Analysen der ELFE-Testergebnisse

Variablenname	Gruppe1	Gruppe2	n1	n2	Statistik	Freiheitsgrade	p	p.adj	cohens d
Gesamt_ELFE	1	2	16	16	-2,019	15	0,031	0,185	-0.505
Gesamt_ELFE	1	3	16	16	-2,267	15	0,019	0,116	-0.567
Gesamt_ELFE	1	4	16	16	-6,904	15	0,00000251	0,0000151*	-1.73
Gesamt_ELFE	2	3	16	16	-0,698	15	0,248	1	-0.174
Gesamt_ELFE	2	4	16	16	-4,507	15	0,000209	0,001*	-1.13
Gesamt_ELFE	3	4	16	16	-4,810	15	0,000115	0,00069*	-1.20

* $p_{adj} < 0.05$

Des Weiteren wurde wie bereits erwähnt eine **Analyse ohne Ausreißer** realisiert, um zu überprüfen, ob diese das Ergebnis verzerren. Hierzu wurden die Personen mit den IDs 110, 120 sowie 123 aus dem Datensatz ausgeschlossen, welcher anschließend 13 Personen enthielt. In Tabelle 10 werden die Ergebnisse der Shapiro-Wilk-Tests dargestellt. Da keiner der Tests signifikant wurde, wird auch hier von einer Normalverteilung ausgegangen.

Tabelle 10 Ergebnisse der Shapiro-Wilk-Tests der ELFE-Testergebnisse - ohne Ausreißer

Messzeitpunkt	Statistik	p
1	0.951	0.615
2	0.961	0.765
3	0.988	0.999
4	0.921	0.258

Nach Berechnung der **rmANOVA** kann ebenfalls ein signifikantes Ergebnis berichtet werden, $F(3, 36) = 16.136$, $p = 0.000000823$, generalisiertes Eta-Quadrat = 0.274. Die Post-hoc Tests ergaben analog zu der Stichprobe mit den Ausreißern signifikante Ergebnisse zwischen dem ersten und dem vierten Messzeitpunkt, $t(12) = -7.11$, $p_{adj} = 0.000037$, $d =$, dem zweiten und dem vierten Messzeitpunkt, $t(12) = -6.16$, $p_{adj} = 0.000146$, $d =$) sowie dem dritten und dem vierten Messzeitpunkt, $t(12) = -4.63$, $p_{adj} = 0.002$, $d =$). In Tabelle 11 werden die Ergebnisse der Post-hoc Analysen der Stichprobe ohne Ausreißer dargestellt.

Tabelle 11 Post-hoc-Analysen der ELFE-Testergebnisse - ohne Ausreißer

Variablenname	Gruppe1	Gruppe2	n1	n2	Statistik	Freiheitsgrade	p	p.adj	Cohens d
Gesamt_ELFE	1	2	13	13	-1,407	12	0,092	0,554	-0.390
Gesamt_ELFE	1	3	13	13	-2,441	12	0,016	0,094	-0.677
Gesamt_ELFE	1	4	13	13	-7,108	12	0,00000617	0,000037*	-1.97
Gesamt_ELFE	2	3	13	13	-1,344	12	0,102	0,612	-0.373
Gesamt_ELFE	2	4	13	13	-6,160	12	0,0000244	0,000146*	-1.71
Gesamt_ELFE	3	4	13	13	-4,632	12	0,000289	0,002*	-1.28

* $p_{adj} < 0.05$

Daraus kann geschlussfolgert werden, dass sich der vierte Erhebungszeitpunkt statistisch signifikant von den vorherigen Zeitpunkten unterscheidet. Hiermit bestätigt sich Hypothese 2.1.3.

4.6.3 Veränderungen in der motivationalen Orientierung

Zur Überprüfung der Hypothesen der Fragestellung 2.2 wurde ein t-Test für abhängige Stichproben durchgeführt. Im Zuge dessen wurde zunächst die **Voraussetzung** der

Normalverteilung der Mittelwertsdifferenzverteilung überprüft. In Tabelle X werden die Ergebnisse des Shapiro-Wilk-Tests der Mittelwertsdifferenzen dargestellt. Die Verteilung der Variable der Vermeidungs-Leistungszielorientierung wird hierbei signifikant, weshalb von einer Verletzung der Normalverteilungsannahme ausgegangen wird und deshalb auf ein nonparametrisches Verfahren für diese Variable zurückgegriffen wird.

Tabelle 12 Ergebnisse des Shapiro-Wilk-Tests der SELLMO-Testergebnisse

	Statistik	p
Lernziel	0.90955	0.08447
Annäherungs-Leistungsziel	0.93065	0.1993
Vermeidungs-Leistungsziel	0.89101	0.04016*
Arbeitsvermeidung	0.97691	0.9127

* $p < 0.05$

Die Ergebnisse der **t-Test** (Lernziel, Annäherungs-Leistungszielorientierung und Arbeitsvermeidung) werden in Tabelle 13 dargestellt. Hierbei wurde keiner der Tests signifikant. Die Effektgröße der Lernzielorientierung beträgt $d = -0.118$, was nach Cohen (1988) einem kleinen Effekt entspricht.

Tabelle 13 Ergebnisse der t-Test für die berechneten parametrischen Tests der SELLMO-Testergebnisse

Variablenname	Gruppe1	Gruppe2	n1	n2	Statistik	df	p	Cohens d
Lernziel	3	4	18	18	-0,499	17	0,312	-0.118
Annäherungs-Leistungsziel	3	4	18	18	0,049	17	0,961	0.0116
Arbeitsvermeidung	3	4	18	18	0,281	17	0,391	0.0663

Der **Wilcoxon-Vorzeichen-Rang-Test**, der aufgrund der nicht-angenommenen Normalverteilung zur Analyse der Vermeidungs-Leistungszielorientierung angewandt wurde, ergab ebenfalls kein statistisch signifikantes Ergebnis, $V = 94$, $p = 0.1868$, $r = 0.329$. Hier handelt es sich nach Cohen (1988) um einen mittleren Effekt.

5. Qualitative Analyse

5.1 Fragestellungen

Im Folgenden sollen die Fragestellungen 1.2 bis 1.10 anhand der Ergebnisse der qualitativen Analyse beantwortet werden. In diesem Zuge wurden, wie bereits beschrieben, qualitative Interviews mit drei Fachkräften, der Programmleitung sowie der Lehrkraft durchgeführt. Die zu beantwortenden Fragestellungen werden allerdings nicht in Kapitel 5 überprüft, sondern in Kapitel 6. Dies hat den Grund, dass in Kapitel 5 lediglich die Ergebnisse der qualitativen Interviews dargestellt werden sollen, die dann in Kapitel 6 im Rahmen der formativen und

summativen Evaluationsziele herangezogen werden, um die Fragestellungen zu beantworten. Des Weiteren wird die qualitative Analyse – im Gegensatz zur quantitativen Analyse – zunächst theoretisch fundiert. Dies hat damit zu tun, dass einerseits diese Arbeit das einzige schriftliche Produkt der darin beschriebenen Evaluation ist und deshalb darauf Wert gelegt wurde, den Programmbeteiligten die für sie unbekannte qualitative Forschungsmethodik zu erläutern, und dass andererseits die qualitative Analyse dem Forschenden große interpretative Freiheiten lässt. Damit die Leserschaft dieser Arbeit die gewonnenen Erkenntnisse der qualitativen Analyse besser nachvollziehen können, sollen deshalb möglichst viele im Laufe des Interpretationsprozesses getroffenen Annahmen expliziert werden.

5.2 Theoretische Grundlagen

5.2.1 Forschungsprogramm Subjektive Theorien

Da qualitative Forschung Aussagen über den Menschen und dessen Zustände trifft, gehen mit jenen Aussagen immer auch bestimmte anthropologische Annahmen einher. Hierbei lässt sich vorstellen, dass jene Annahmen auch bei Wissenschaftlerinnen und Wissenschaftlern bewusst und unbewusst vorhanden sind und den Forschungsprozess und dessen Reflexion beeinflussen. Um einerseits größtmögliche Transparenz und intersubjektive Nachvollziehbarkeit zu gewährleisten und andererseits den eigenen Denkprozess als forschende Person zu strukturieren, sollen diese Annahmen im Folgenden kurz und bündig expliziert werden, ohne inhaltlich zu sehr auszuschweifen. In diesem Rahmen soll der Forschungsgegenstand der qualitativen Analyse klar und präzise formuliert werden. Gleichzeitig soll, wie bei Mayring (2022) beschrieben, theoretisch über die Entstehungsbedingungen des Sinnes reflektiert werden, um diese als darauf aufbauenden Schritt in der Analyse des Datenmaterials theoretisch orientiert zu reflektieren.

Diese Arbeit greift auf das Forschungsprogramm Subjektive Theorien zurück, wobei dies eine selektive Wahl aus mehreren Alternativen ist (Groeben et al., 1988). In diesem Zuge werden einerseits die anthropologischen Annahmen für diese Arbeit übernommen und andererseits wird das Konzept "Subjektive Theorien" als analytische Einheit übernommen, mit welcher das im Rahmen des qualitativen Forschungsprozesses gewonnene Material analysiert werden kann. Die inhaltliche Adäquatheit des Forschungsprogramms Subjektive Theorien für die in Kapitel 5.1 formulierten Fragestellungen soll im Folgenden ebenfalls erläutert werden.

Groeben und Scheele (2010) wenden dasselbe Menschenbild auf Forschungssubjekt und Forschungsobjekt an, das bedeutet, die anthropologischen Annahmen gelten auch für die Wissenschaftlerin und den Wissenschaftler. Daraus folgt die zentrale Annahme, dass der Mensch zur (Selbst-)Erkenntnis fähig ist, in anderen Worten: Der Mensch kann Wissen über sich selbst und andere Mensch oder einen Gegenstand gewinnen ("epistemologische Subjektmodell") (Groeben et al., 1988). Auf dessen Grundlage treffen sie noch weitere, darüber hinaus gehende Annahmen: Der Mensch ist sprach- und kommunikationsfähig, er ist reflexiv, er besitzt die Fähigkeit zum (bewussten und intentionalen) Handeln und zur Rationalität (vgl. Groeben, 1986). Ausgehend von der Annahme der Intentionalität beschreiben Groebe und Scheele (2010) den Zusammenhang zwischen Handeln und Sinn, wobei der Sinn analog als das Ziel der Handlung verstanden werden kann. Hierdurch wird die Schnittstelle zur qualitativen Forschung geschaffen, deren Ziel das Verstehen des

subjektiven Sinnes ist (z. B. Lamnek, 2005). Das bedeutet, im Kontext des Forschungsprogramms Subjektive Theorien und eines qualitativen Zugangs handelt es sich bei dem Erkenntnisgegenstand um den Sinn, den ein Subjekt mit seinen Handlungen verbindet. Handlungen beschreiben allerdings nur einen Teilbereich des menschlichen Verhaltens, nämlich jenes Verhalten, welches intentional erfolgt (vgl. Groeben, 1986). Hierbei nimmt man an, dass das Handeln situiert ist und mehrere Handlungsalternativen bestehen, aus denen das Individuum wählen kann. Handeln ist somit eine reflexive Entscheidung, eine bestimmte Aktion zu realisieren und die Aktion zu planen und zu kontrollieren (Groeben & Scheele, 2010). Relevante Einflussfaktoren auf das Verhalten sind demnach die situativen Kontextfaktoren und individuelle und soziale Normen- und Wertesysteme (Groeben & Scheele, 2010).

Der Prozess, der den Sinn mit einer Handlung verbindet und eine Handlung aus verschiedenen Alternativen auswählt, ist auf der kognitiven Ebene zu verorten. Das Forschungsprogramm Subjektive Theorien definiert hierbei Kognitionen in einem anderen Sinne als in der traditionellen Psychologie üblich (vgl. Groeben & Scheele, 2010). Kognitionen beziehen sich zwar analog auf mentale Prozesse, allerdings werden diese in niedrigkomplexe Kognitionen (Begriffe, Sätze) und hochkomplexen Kognitionen (Aggregate von Begriffen und Sätzen) strukturiert (Groeben & Scheele, 2010). Subjektive Theorien lassen sich demzufolge auf der Ebene der hochkomplexen Kognitionen verorten und verknüpfen Handlung mit Sinn. Der Begriff Subjektive Theorie soll hier explizit eine Analogie zur wissenschaftlichen Theorie ziehen. Wissenschaftliche Theorien dienen nach Groebe et al. (1988) der Erklärung, Prognose und Technologie in Bezug auf ihren Gegenstand. Plakativ lassen sich diese Funktionen wie folgt erläutern: *Erklärung* bezeichnen gegenstandsbezogene Aussagen über Ursache-Wirkungs-Zusammenhänge. *Prognosen* sagen ausgehend von einem Referenzzustand einen zukünftigen Zustand vorher. Und *Technologien* beschreiben konkretes Handlungswissen, wie ein bestimmter Zustand hergestellt werden kann. Subjektive Theorien erfüllen demnach dieselben Funktionen wie wissenschaftliche Theorien, allerdings müssen sie nicht den wissenschaftlichen Kriterien gerecht werden. So werden bei subjektiven Theorien die subjektiven Hypothesen nicht systematisch, explizit und präzise aus den Theorien abgeleitet – wie bei wissenschaftlichen Theorien – sondern es herrscht eher eine (implizite) argumentative Verknüpfung zwischen beiden vor (Groebe & Scheele, 2010). Wagner (2016) macht darauf aufmerksam, dass sich wissenschaftliche Theorien insbesondere dadurch auszeichnen, dass sie durch die Wissenschaftstheorie fundiert werden, welche die erkenntnistheoretischen und begrifflichen Voraussetzungen des wissenschaftlichen Erkenntnisprozesses systematisch reflektiert und hinterfragt. Begriffe, die in wissenschaftlichen Theorien verwendet werden, müssen des Weiteren präzise definiert werden. Darüber hinaus werden die Theorien systematisch und kritisch (an der Realität) überprüft (Wagner, 2016). All diese Kriterien scheinen keine notwendigen Bedingungen für subjektive Theorien zu sein. Allerdings sollte der streng wissenschaftliche Charakter von wissenschaftlichen Theorien nicht überinterpretiert werden, da sie nach Kuhn (1962) immer auch Ergebnis eines sozialen Konsenses sind, der sich abhängig der dominierenden sozialen Gruppe innerhalb des Wissenschaftssystems auch ändern kann („wissenschaftliche Revolution“). Somit unterliegen wissenschaftliche Theorien auch sozialen Einflussbedingungen und nicht nur rein rationalen Kriterien.

Ausgehend von diesen Voraussetzungen definieren Groeben et al. (1988, S. 19) subjektive Theorien als Kognitionen der Welt- und Selbstsicht, welche aus komplexen Aggregaten von Begriffen und Sätzen bestehen, die (implizit) argumentativ strukturiert sind.

Sie erfüllen dieselben Funktionen wie wissenschaftliche Theorien (Erklärung, Prognose und Technologie). Anhand dieser Definition lässt sich der Begriff des Epistemologischen Subjektmodells, der die anthropologischen Annahmen des Forschungsprogramms Subjektive Theorien beschreibt, begrifflich besser fassen. Der Mensch ist somit ein Subjektiver Theoretiker, der seine (intendierten) Handlungen mithilfe von Theorien fundiert. Das bedeutet beispielsweise: Handlungen werden realisiert, da man diese mit einem bestimmten Ziel (Sinn) in Beziehung setzt, wobei sich diese Beziehung aus der subjektiven Theorie ergibt, die beispielsweise aus einer Erklärung des handlungsbezogenen Ursache-Wirkungs-Zusammenhangs besteht. Konsequenterweise bedeutet dies, dass nach dem Forschungsprogramm Subjektive Theorien der subjektive Sinn (intendierter Handlungen) immer auch mit (impliziten) subjektiven Theorien in Verbindung steht. Würde man nun menschliches Verhalten von außen beobachten, hätte man zunächst keine valide Informationsgrundlage hinsichtlich deren subjektiven Sinnes oder Theorien. Diese sind alleinig durch die Innenansicht des Subjekts beobachtbar. Hierbei wird angenommen, dass sich diese Innenansicht (sprachlich) kommunikativ mitteilen lässt und die subjektiv-theoretischen Reflexionen somit (sprachlich) expliziert werden können (vgl. Groeben & Scheele, 2010). Um somit Zugang zu dem subjektiven Sinn und Theorien zu bekommen, bedarf es der Herstellung von Situationen, in denen der Forschende dem Befragten gezielt zu interessierenden Themen Fragen stellen und der Befragte dem Forschenden sprachlich-kommunikativ Auskunft in Bezug auf das adressierte Thema geben kann (Interviewsituation). Hier lassen sich allerdings eine Einschränkung skizzieren: Das Individuum kann sprachlich eine falsche Aussage treffen hinsichtlich der Attribuierung eines Sinns zu seiner Handlung (Groeben & Scheele, 2010). Dies kann verschiedene Gründe haben: Es kann sich einerseits dabei irren, indem sich die eigenen Motive nicht eingestanden werden oder aber es kann aufgrund von sozialer Erwünschtheit den Sinn verfälscht oder gar nicht darstellen (Groeben & Scheele, 2010). Des Weiteren lässt es sich vorstellen, dass oftmals Informationen fehlen, um vollständig Auskunft über die subjektive Theorie des Befragten zu erlangen. Gleichzeitig kann dafür argumentiert werden, dass sich der Sinn durch unbewusste Mechanismen konstituiert, die dem Individuum bewusst (ad hoc) nicht wahrnehmbar und somit kommunizierbar sind. Dies widerspricht allerdings der im Rahmen dieser Arbeit getroffenen Annahme, dass Handlungen intendiert und bewusst sind. Aufgrund dessen fokussiert sich diese Arbeit vornehmlich auf bewusste Sinndimensionen.

Da in dieser Arbeit der Sinn analog als Ziel einer Handlung aufgefasst wird, ergibt sich durch die Betrachtung der zugrunde liegenden subjektiven Theorien eine weitere Analyseebene, die eine Erklärung zulässt, welche möglichen Ziele (Sinne) die Akteurinnen und Akteure in der Praxis ausgehend von diesen subjektiven Theorien verfolgen könnten. Der Rückgriff auf das Forschungsprogramm Subjektive Theorien im Kontext der Explizierung des Forschungsgegenstandes bringt somit den Vorteil mit sich, den erfragten Sinn um subjektiv-theoretische Annahmen zu erweitern. Es lässt sich vorstellen, dass durch die Explizierung der subjektiven Theorien den Akteurinnen und Akteuren in der Praxis bei Übereinstimmung und Nachvollziehbarkeit eine weitere Reflexionsebene eröffnet werden könnte, die einen Anknüpfungspunkt schafft, diese subjektiven Theorien den wissenschaftlichen (objektiven) Theorien gegenüberzustellen. Straub und Weidemann (2015) argumentieren, dass durch diese Gegenüberstellung eine Modifikation der subjektiven Theorien zugunsten der wissenschaftlichen Theorien bewirkt werden kann. Dies kommt dem formativen Evaluationsziel der Evidenzbasierung des Programms und der Professionalisierung dessen Mitarbeitenden zugute. Einschränkend lässt sich hierbei aber anmerken, dass durch den

Fokus auf subjektive Theorien rein kognitive Merkmale bzw. Produkte in dieser Evaluationsstudie betrachtet und motivationale oder emotionale Merkmale eher ausgeklammert werden.

Zusammenfassend lässt sich sagen, dass sich durch das Forschungsprogramm Subjektive Theorien einerseits ein für die in dieser Arbeit adressierten Evaluationsfragestellungen interessanter analytischer Zugang zu den subjektiven Sinndimensionen der Befragten ergibt, der allerdings auch bestimmte Annahmen und Limitationen mit sich bringt. Somit wird es möglich, konkret zu reflektieren, welche theoretischen Annahmen hinter den Aussagen der Befragten stehen könnten. Dies erscheint insbesondere für die Fragestellung 1.7 interessant, um zu eruieren, welche theoretische Annahmen über Wirkungen von pädagogischen Handlungen beobachtbar sind. Wie bereits ersichtlich wurde, greift diese Arbeit auf Interviews mit dem pädagogischen Fachpersonal von ENB und der Lehrkraft zurück, um Zugang zu deren subjektiven Sinndimensionen und Theorien zu erlangen. Um dem Anspruch der Wissenschaftlichkeit dieser Evaluationsstudie gerecht zu werden, wird das dadurch gewonnene Datenmaterial mittels der Qualitativen Inhaltsanalyse nach Mayring ausgewertet, welche im folgenden Kapitel theoretisch erläutert wird.

5.2.2 Auswertung Qualitative Inhaltsanalyse nach Mayring

Ziel von Mayring (2022) ist es, eine systematische interpretative Methode zu beschreiben, mit welcher latente Sinnstrukturen in sprachlichem Material systematisch erschlossen werden können. Diese Methode soll zumindest dem wissenschaftlichem (quantitativem) Gütekriterium der Objektivität (als intersubjektive Nachvollziehbarkeit) gerecht werden, um der quantitativen Forschung eine systematische und objektive qualitative Auswertungsmethode gegenüberzustellen. Bevor detailliert beschrieben werden soll, wie die Qualitative Inhaltsanalyse zu realisieren ist, wird zunächst deren Gegenstand und an sie gestellte Anforderungen dargeboten (Mayring, 2022): Der Gegenstand der Qualitativen Inhaltsanalyse ist Material, das aus irgendeiner Art von Kommunikation stammt. Kommunikation meint hier die Übertragung von Symbolen mittels verschiedener Medien, wie Sprache, Bilder, Gesten etc. Diese Kommunikation muss in fixierter, dokumentierter Form vorliegen. Zur Auswertung des Materials muss nun systematisch – im Sinne von einem vordefinierten System folgend – vorgegangen werden, um insbesondere das Vorgehen des Forschenden unter vorher festgelegten Regeln zu subsumieren. Hierdurch wird die intersubjektive Nachvollziehbarkeit gewahrt. Eine dieser bei der Auswertung zu befolgende Regel ist die Theoriegeleitetheit. Das bedeutet einerseits, dass sich das Vorgehen der Auswertung an einer (wissenschaftlichen) Theorie orientiert; hier: Die Theorie der Qualitativen Inhaltsanalyse nach Mayring. Andererseits bedeutet es auch, dass die Fragestellung mithilfe von theoriegeleiteten Interpretationen anhand des Datenmaterials zu beantworten versucht wird. Theorien definiert Mayring (2022) in diesem Kontext als System allgemeiner Sätze über den zu untersuchenden Gegenstand: Diese Sätze beschreiben die Erfahrungen anderer Forschenden, die diese in ein Aussagensystem zusammengefasst und systematisiert haben. Theoriegeleitetheit heißt nun, an diese Erfahrungen anzuknüpfen, um einen Erkenntnisfortschritt zu erreichen.

Damit nun die Inhaltsanalyse „qualitativ“ wird und somit die (qualitativ ausgerichtete) Fragestellung beantwortet werden kann, muss aus ihr das Rekonstruieren des qualitativen Sinnes erfolgen können. In diesem Sinne beschreibt Mayring (2022) die Inhaltsanalyse als

schlussfolgernde Methode, die nicht nur den „Inhalt“ des Materials analysiert, sondern darüber hinaus Rückschlüssen auf die Kommunikation und die Kommunikationssituation zieht. Aus der Erschließung dieser Kommunikationsinhalte und –merkmale werden laut Mayring (2022) jene qualitativen, sinnbezogenen Aussagen möglich. Ein zentrales Element der Inhaltsanalyse stellen in diesem Sinne Kategorien dar; allerdings definiert Mayring (2022) den Begriff der Kategorie nicht. Anstatt dessen scheint es so, als lehne er sich an der Definition von Kategorie aus der Quantitativen Inhaltsanalyse an (Mayring, 2022): Kategorien sind Begriffe, das bedeutet, sie haben eine extensionale und eine intensionale Dimension. Darüber hinaus wird implizit angenommen, dass mindestens ein anderer Begriff existiert, der sich unter diese (begriffliche) Kategorie subsumieren lässt und somit extensional unter diesen Begriff fällt. Des Weiteren haben Kategorien bei Mayring (2022) eine Doppelfunktion: Einerseits setzt sich der Forschende eine „Kategorienbrille“ auf, mit der er das Datenmaterial analysiert und andererseits ist das Produkt der Qualitativen Inhaltsanalyse ein Kategoriensystem, welches sich aus dem Datenmaterial ergeben hat. Ersteres wird als deduktive Kategorienbildung bezeichnet, bei welcher der Forschende (theoretisch abgeleitete) Kategorien auf das Datenmaterial anwendet. Zweiteres wird als induktive Kategorienbildung aufgefasst, in dessen Zuge die Kategorien erst durch Abstraktion des Datenmaterials geschaffen werden. Ziel des systematischen Vorgehens im Rahmen der Qualitativen Inhaltsanalyse ist somit das Schaffen eines Kategoriensystems. Das Kategoriensystem soll alle (interessierenden) Aspekte eines Textes (erschöpfend) erfassen. Hierbei werden konkrete Textstellen systematisch den Kategorien im Kategoriensystem zugewiesen. Gleichzeitig können Kategorien aus beliebig vielen Unterkategorien bestehen, wodurch eine hierarchische Kategorienstruktur entsteht

Der Rückbezug zur Qualitativen Forschung besteht nun darin, dass Qualitative Forschung immer begrifflich denkt. Der Forschungsgegenstand wird begrifflich beschrieben und gleichzeitig befinden sich die sprachlichen Äußerungen des Befragten auf begrifflicher Ebene. Durch den Fokus der Qualitativen Inhaltsanalyse auf (begriffliche) Kategorien und deren systematischen Zusammenhänge wird nun der Denkprozess des Forschenden strukturiert, systematisiert und somit intersubjektiv nachvollziehbar. Um nun systematisch von dem Datenmaterial ausgehend auf latente Sinngehalte schließen zu können und das Kategoriensystem zu erstellen, beschreibt Mayring (2022) drei voneinander unabhängige Grundtechniken der Qualitativen Inhaltsanalyse: Die Zusammenfassung, die Explikation und die Strukturierung.

Zusammenfassung:

Die Zusammenfassung beschreibt den Prozess der Abstraktion des vorliegenden Datenmaterials, dergestalt, dass es immer noch ein Abbild des Grundmaterials ist (Mayring, 2022). Hierbei werden bestimmte Bestandteile des Textes nach einem bestimmten Definitionskriterium abstrahiert und andere exkludiert. Darunter lässt sich analog ebenfalls die induktive Kategorienbildung subsumieren.

Explikation:

Explikation bedeutet, dass einzelne, interessant erscheinende Textteile aus dem Datenmaterial genau untersucht werden. Zu deren Untersuchung zieht man zusätzliches Material heran, das sich eignet, die Textstelle zu „explizieren“ bzw. das eigene Verständnis zu erweitern.

Strukturierung:

Mittels der Strukturierung werden bestimmte Aspekte aus dem Material selektiert, um so die innere Struktur des Materials zu extrahieren. Hierbei lassen sich ebenfalls externe Kriterien zur Strukturierung des Textes heranziehen (deduktive Kategorienbildung).

Diese drei Grundtechniken strukturieren somit den Auswertungsprozess. Da es neben dem inhaltsanalytischen Ansatz von Mayring noch weitere Ansätze gibt, bedarf es bestimmter Entscheidungskriterien, um eine Auswahl zu begründen. In dieser Arbeit wurde sich an Mayring orientiert, um die in den Aushandlungsprozessen festgelegten Evaluationsfragestellungen und deren theoretische Fundierung als Kategorien in die Analyse zu integrieren. Dies hat den Grund, dass Mayring insbesondere die deduktive Kategorienbildung an den Anfang des inhaltsanalytischen Prozesses situiert. Um den Auswertungsprozess eine übergeordnete Struktur zu geben, wurde sich an den Best-Practice-Guidelines von Clarke und Braun (2013) orientiert, die den Prozess der Entwicklung von Interviewfragen in sechs Phasen kategorisieren:

1. Zunächst soll sich der Forschende mit den Daten vertraut machen.
2. Anschließend sollen (induktiv und deduktiv) anfängliche Kategorien geschaffen werden.
3. Dann soll explizit und systematisch im Material nach Kategorien gesucht werden.
4. Die gefundenen Kategorien sollen nun definiert und benannt werden.
5. Darauffolgend soll sich noch ein Überblick über die gefundenen Kategorien geschaffen werden.
6. Abschließend schlagen sie vor, den Bericht zu schreiben.

5.3 Empirische Untersuchung

5.3.1 Erhebungsmethode und Stichprobe

Da im Rahmen des Evaluationsprozesses zunächst ausgehandelt werden musste, welche Ressourcen seitens ENB für die Evaluation zur Verfügung gestellt werden konnten, ist die Erhebungsmethode und die Stichprobe Produkt dieser Aushandlung. In Übereinstimmung mit dem in Kapitel 5.2.1 dargestellten Forschungsprogramm Subjektive Theorien wurden Einzel-Interviews realisiert, um systematische Kommunikationssituationen herzustellen. Das bedeutet, dass das Interview nicht in Gruppensituation stattfand, sondern als Einzelgespräche zwischen dem Autor dieser Studie und den Befragten. Hierfür wurden als Resultat des Aushandlungsprozesses mit ENB festgelegt, dass die Interviews ca. 30 Minuten dauern sollten.

Innerhalb der Interviews wurde in einem Abschnitt die „Critical Incident Technique“ (CIT) in modifizierter Form angewandt. Ziel ist das Aufdecken von Verhaltensmustern, die zur Bewältigung von kritischen Situationen eingesetzt werden (Reuschenbach, 2008). Kritische Situationen beschreiben hierbei herausfordernde und teils unerwartete Situationen (vgl. Nixdorf, 2020). Gleichzeitig sind solche Situationen „kritisch“ für den Erfolg oder Misserfolg eines Prozesses. Diese Arbeit orientiert sich an der Definition von Hettlage und Steinlin (2006): Sie fassen CIT als qualitatives Tool zur Analyse von Problemlöseverhalten auf. Hierbei wird Bezug auf eine kritische Problemsituation genommen und in eine kurze Rahmengeschichte gepackt. Hierdurch werden laut Hettlage und Steinlin (2006) Reflexionsprozesse bei den Befragten ausgelöst und animieren deren individuelles

Problemlöseverhalten. Aufgrund der geringen Zeitspanne des Interviews nahmen die CIT-Interviews lediglich fünf bis zehn Minuten der gesamten Interviewdauer ein. Dadurch bedingt konnte dem Anspruch, dass die Befragten möglichst detailliert die eigenen Gedanken und Erfahrungen schildern, nicht gerecht werden. In diesem Zuge wurden, wie bereits beschrieben, drei pädagogische Fachkräfte, die Programmleitung und die Klassenlehrkraft interviewt.

5.3.2 Datenerhebung

Die Interviews wurden in der 16. Kalenderwoche in 2024 (April) geführt. Mit der Programmleitung wurde in der 20. Kalenderwoche in 2024 (Mai) ein weiteres Interview durchgeführt. Die Vermittlung der jeweiligen Interviewzeitpunkte geschah mittels Email- und telefonischen Austausch zwischen dem Autor dieser Studie und der Programmleitung. In Tabelle 14 wird die jeweilige Dauer sowie die insgesamt Durchschnittsdauer der qualitativen Interviews dargestellt.

Tabelle 14 Dauer der qualitativen Interviews in Minuten

Interviewnummer	1	2	3	4	5	6	Durchschnitt
Dauer in Minuten	34	39	33	40	37	64	41,17

Interviewort war die staatliche Grundschule an der Walliser Straße, an der auch das Förderprogramm von ENB stattfindet. Die Interviews wurden mit der jeweiligen Person im Anschluss an deren Programmstunde an der Schule realisiert, wodurch diese an verschiedenen Wochentagen zu unterschiedlichen Zeiten stattfanden. Im Schulgebäude fanden einerseits zwei Interviews im Werkraum statt, in dem das Programm einen Teil seiner Förderstunden abhält, und andererseits in einem Differenzierungsraum, in dem normalerweise die individuelle Förderung stattfindet. Die Wahl der Räume geschah pragmatisch und situationsabhängig. Ziel der Situierung im Schulgebäude war es, die Interviews in für die Befragten vertrauten Räumlichkeiten zu realisieren.

5.3.3 Interviewleitfaden & Forschungsethik

Um die Interviews zu strukturieren und systematisch auf die Fragestellungen zu beziehen, wird ein Leitfaden entworfen. Strübing (2018) beschreibt einen Leitfaden als Orientierungshilfe und nicht als konkreter Ablaufplan, der streng befolgt werden muss. Diese Arbeit orientiert sich vornehmlich an Helfferich (2011), welcher einen konkreten Vorschlag zur Leitfadenkonstruktion unterbreitet. Die in dieser Arbeit entwickelten Leitfäden sind im Anhang aufgeführt. Helfferich (2011) definiert einen Leitfaden als ein Instrument, in welchem Anweisungen für Fragen und Erzählaufforderungen festgeschrieben werden. Allerdings soll der Leitfaden keine bloße Liste an Fragen sein. Fragen sollen anstatt dessen strukturiert und in eine Reihenfolge gebracht werden sowie nicht nur inhaltsbezogen sein, sondern auch zur Steuerung und Aufrechterhaltung des Gesprächs dienen. Die Themen sollten so strukturiert werden, dass es zu keinen abrupten Sprüngen und Themenwechseln kommt (vgl. Helfferich, 2011). Ein Leitfaden soll somit eine gewisse Struktur für das Interview vorgeben, gleichzeitig allerdings maximale Offenheit zulassen. Dies geschieht, indem lediglich der Erzählraum durch

den Leitfaden thematisch strukturiert wird und teilweise einen spezifischen Fokus legt. Denn: Im Interview sollte laut Helfferich (2011) immer die spontan produzierte Erzählung im Vordergrund stehen. In diesem Zuge gibt Helfferich (2011) einige Empfehlungen für die Leitfadenskonstruktion: Einerseits soll es im Leitfaden maximal vier Blöcke geben, die durch Erzählaufforderungen eingeleitet werden und die hierarchisch geordnete Fragen enthalten, die flexibel gestellt werden können. Andererseits empfiehlt er, am Ende des Leitfadens noch eine Frage einzubauen, die dem Befragten die Möglichkeit eröffnet, eigene relevante Themen oder nicht genannte Sachverhalte zu erwähnen. Für das Erstellen des Leitfadens entwickelte Helfferich (2011) die SPSS-Methode, an der sich in dieser Arbeit ebenfalls orientiert wurde. Diese gibt konkrete Arbeitsschritte vor, die den Konzeptionsprozess strukturieren. SPSS steht für Sammeln, Prüfen, Sortieren und Subsumieren.

- **Sammeln:** Im ersten Schritt "Sammeln" sollen alle möglichen Fragen gesammelt werden, die bzgl. des Forschungsgegenstandes interessant erscheinen. Ziel dabei ist es, möglichst viele Fragen ohne jegliche Selektionskriterien zunächst zu sammeln.
- **Prüfen:** Anschließend wird beim "Prüfen" dieser Fragenkatalog reduziert und strukturiert, indem Fragen exkludiert oder paraphrasiert werden. Helfferich (2011) nennt hier verschiedene Kriterien, die angewandt werden können, wie den Ausschluss aller Faktenfragen (sollen laut ihm durch einen Fragebogen erhoben werden), der Ausschluss redundanter Fragen sowie den Ausschluss derjenigen Fragen, die lediglich die eigenen Annahmen bestätigen sollen.
- **Sortieren:** Der dritte Schritt "Sortieren" beschreibt die Kategorisierung der Fragen in (vier) thematische Bündel.
- **Subsumieren:** Beim "Subsumieren" werden nun abschließend für die einzelnen Bündel möglichst einfache Erzählaufforderungen formuliert, unter die die selektierten Fragen im zweiten Schritt subsumiert werden. Anforderung an den Erzählimpuls ist, dass dieser eine Erzählung evoziert, die möglichst viele der interessierenden Aspekte tangiert.

Des Weiteren legt Helfferich (2011) einen besonderen Schwerpunkt auf forschungsethische Aspekte bei der Gestaltung des Leitfadens. Einerseits sollte forschungsethisch die informierte Einwilligung der Befragten vor Durchführung des Interviews eingeholt werden sowie die Nicht-Schädigung der Teilnehmenden durch die Interviewteilnahme gewährleistet werden. Dies meint bspw., dass den Teilnehmenden durch die Teilnahme (aber auch der Nicht-Teilnahme) am Interview keine Nachteile entstehen. Die informierte Einwilligung bezieht sich insbesondere darauf, dass der vorgesehene Forschungszweck sowie die Träger, Verantwortlichen und Kooperationspartner der Studie kommuniziert werden. Ein weiterer, zentraler forschungsethischer Aspekt ist die Anonymisierung der Daten. Um eine möglichst offene Gesprächsatmosphäre zu schaffen und dem forschungsethischen Anspruch gerecht zu werden, keine Nachteile durch die Interviewteilnahme und der darin getätigten Äußerungen zu generieren, wird in dieser Arbeit auf personenbezogene Beschreibungen verzichtet. Wie bereits dargestellt, werden lediglich im Rahmen der Stichprobenbeschreibung zentrale personenbezogene Eckdaten dargestellt. In der Darstellung der Ergebnisse und der Interpretation der qualitativ untersuchten Daten werden somit keine personenbezogenen Informationen oder Interpretationen geäußert. Darüber hinaus bedeutet dies, dass die Interviewtranskripte nicht veröffentlicht werden und somit nicht im Rahmen dieser Studie dargestellt werden können. Helfferich (2011) führt noch weitere Kriterien auf, an denen sich in dieser Arbeit orientiert wurde und die in der Einwilligungserklärung (im Anhang aufgeführt) zum Tragen kamen. Diese Einwilligungserklärung wurde jeweils vor Durchführung der

Interviews den Befragten vorgelegt und von ihnen unterzeichnet. Da nun beschrieben wurde, nach welchen Kriterien der Leitfaden entwickelt wurde, wird im Folgenden erläutert, welche Kategorien in den Leitfaden aufgenommen wurden und wie diese Kategorien zustande kamen.

5.3.4 Festlegen der Kategorien für den Leitfaden

Die Kategorien im Leitfaden dienen der thematischen Gliederung des Leitfadens. Sie bestimmen somit die von Helfferich (2011) beschriebenen thematischen Bündel. In dieser Arbeit wurde sich an der Empfehlung Helfferichs (2011) orientiert, lediglich vier übergeordnete thematische Bündel für den Leitfaden heranzuziehen. Grundlage der Festlegung der Kategorien war der in Kapitel 3.3 beschriebene Aushandlungsprozess, in welchem u. a. die Evaluationsfragestellungen festgelegt wurden (siehe Anhang für das entsprechende Dokument). Auf Grundlage dieser ursprünglichen Ausarbeitung und weiterer Recherchen seitens des Autors dieser Studie wurden – wie in Kapitel 3.3 beschrieben – die thematischen Bündel festgelegt. Bei der Leitfadenskonzepion wurden zwei leicht differierende Leitfäden entwickelt, um die unterschiedlichen Perspektiven des pädagogischen Fachpersonals seitens ENB einerseits und die der Klassenlehrkraft andererseits einzufangen. In diesem Zuge wurde somit ein Leitfaden für die vier Mitarbeiter von ENB und ein Leitfaden für die Klassenlehrkraft konzipiert. Da mit der Programmleitung ein zweites Interview zu einem späteren Zeitpunkt durchgeführt wurde, wurde hierfür ebenfalls ein weiterer Leitfaden erstellt. Resümierend bedeutet dies, dass insgesamt sechs Interviews durchgeführt und hierfür drei unterschiedliche Leitfäden erstellt wurden. All diese Leitfäden sind im Anhang ersichtlich.

Unabhängig des Leitfadens wurde zunächst vor Beginn des Interviews die Einwilligungserklärung dem Befragten vorgestellt und erläutert und anschließend von diesem unterzeichnet. In diesem Zuge stellte sich die Person des Autors dieser Studie noch einmal kurz vor und erläuterte die Evaluationsziele. Anschließend wurde die Aufnahme des Interviews gestartet.

Leitfaden Personal:

Zunächst sollen diejenigen zwei Leitfäden erläutert werden, die in der 16. Kalenderwoche stattfanden. Der Leitfaden für das spätere Interview mit der Programmleitung soll im Anschluss kurz dargestellt werden. Da sich der Leitfaden der Lehrkraft in einigen Punkten unterscheidet, soll in der folgenden Beschreibung der thematischen Bündel diese Unterschiede kurz dargestellt werden. Zur genaueren Betrachtung wird auf die Leitfäden in den gerade beschriebenen Anhängen verwiesen. Zum Einstieg in das Interview wurde der oder die Interviewte befragt, ob er oder sie sich kurz vorstellen und seine oder ihre Aufgaben und Rolle in ENB beschreiben könne. Ziel dieser Frage war, einen Redefluss herzustellen und einen leichten Einstieg in das Interview zu ermöglichen. Das erste thematische Bündel wurde im Leitfaden als *Unterrichtskonzeption* beschrieben. Hierbei wurde nach dem typischen Entscheidungsprozess hinsichtlich der Konzeption der Programmangebote gefragt. Dieses thematische Bündel wurde nicht in den Leitfaden der Lehrkraft aufgenommen. Das zweite thematische Bündel wurde als *„Qualität“* bezeichnet. In diesem Zuge wurde einerseits das Verständnis von *„Bildungsqualität“* befragt. Andererseits wurde gefragt, wie und nach welchen Kriterien konkrete pädagogische Handlungen als wirksam beschrieben werden können - vor dem Hintergrund des zuvor geäußerten Qualitätsverständnisses. Dieses thematische Bündel wurde identisch in den Leitfaden der Lehrkraft aufgenommen. Lediglich wurde hier die Frage

ergänzt, wie die Lehrkraft das Qualitätsverständnis von ENB beschreiben würde. Das dritte thematische Bündel befasste sich mit Fragen der *“Evidenzbasierung”*. Hier wurde insbesondere gefragt, inwiefern wissenschaftliche Evidenzen in den pädagogischen Alltag und dort situierten Entscheidungen nach Meinung der Befragten integriert werden könnten. Anschließend wurde das in Kapitel 5.3.1 beschriebene CIT-Interview als eigener Abschnitt innerhalb des Interviews eingeplant.

Hierbei wurde ein *“Schülerprofil”* erstellt, welches im Anhang ersichtlich ist. Dieses Schülerprofil sollte bestimmte quantitative Daten, die im Laufe des Programms über eine anonymisierte Schülerin erhoben wurden, kompakt darstellen. Der Fokus des Schülerprofils wurde auf den sprachlichen Bereich gesetzt, das bedeutet, es wurde ein Teil der für die Sprachkompetenz relevanten Daten dargestellt. Hierzu wurden zunächst einige Schülerdaten (Name, Geschlecht, gesprochene Sprache zuhause, Wiederholung 1. Klasse, Einzelförderung) dargeboten, gefolgt von den Deutschnoten von Klasse 1-3 jeweils aus den Jahreszeugnissen. Hierbei ist anzumerken, dass es bei der Konstruktion des Schülerprofils zu einem Fehler kam. Hierbei wurde irrtümlicherweise angenommen, dass die vom Programm zur Verfügung gestellten Noten der einzelnen Schülerinnen und Schülern die Noten der Klassenstufen 1-3 beschrieben. Allerdings wurden in der ersten Klassenstufe keine Noten vergeben und somit stammten die verwendeten Noten tatsächlich aus den Klassenstufen 2-4, wobei die Note der vierten Klasse aus dem Zwischenzeugnis entnommen wurde. Dieser Fehler konnte erst zu einem späteren, den Interviews nachgelagertem Zeitpunkt, aufgedeckt werden. Für die Aussagekraft der Interpretationen der Interviews wirkt dies allerdings nicht beeinträchtigend, sondern erlaubte es, einen Referenzwert für die erste Klassenstufe darzustellen, der somit der Interpretation zugänglich wurde. Anschließend wurde die durch den SELLMO-Fragebogen erhobene Lernmotivation aus der 3. Klasse sowie das durch den ELFE-Fragebogen erhobene Wortverständnis von Klasse 1-3 dargestellt. Sowohl die Noten als auch die Lernmotivation und das Wortverständnis wurden jeweils zusammen mit dem Durchschnittswert der Klasse abgebildet, um direkte Vergleiche zwischen der anonymen Schülerin und der Klasse zuzulassen. Dieses Schülerprofil wurde mittels der CIT in das Interview integriert. Hierbei wurde den Befragten zunächst erläutert, dass ein Ziel der (formativen) Evaluation die Eruiierung der Nutzung der quantitativ erhobenen Daten in alltäglichen pädagogischen Entscheidungen ist und nicht lediglich das Heranziehen der quantitativen Daten zur Kontrolle der Wirkung des Programms. Anschließend wurde das Schülerprofil vorgestellt und Schritt für Schritt erläutert. Daraufhin wurden die Befragten aufgefordert, das Schülerprofil zu interpretieren und auch bzgl. seiner Praxistauglichkeit zu bewerten. Dieser Teil des Leitfadens wurde weniger strukturiert und stärker offengehalten, um den Befragten möglichst viel Freiraum zu lassen. Das CIT-Interview wurde identisch in den Leitfaden der Lehrkraft eingebunden. Abschließend wurden Einstellungsfragen in Bezug auf das *Programm* ENB gestellt. Dabei wurde gefragt, was die befragte Person *“allgemein”* vom Programm hält, um möglichst individuelle Schwerpunktsetzungen zuzulassen. Zum Teil wurden an dieser Stelle Einschätzungen zur Wirksamkeit des Programms befragt, falls dies nicht an einer vorherigen Stelle des Interviews aufkam. Dieser Teil wurde ebenfalls in den Leitfaden der Lehrkraft integriert, allerdings wurde er zum Einstieg in das Interview gestellt, um den Fokus der Befragung gezielt auf ENB zu legen.

Erweiterter Leitfaden für Programmleitung:

Wie bereits beschrieben, fand das zweite Interview mit der Programmleitung vier Wochen nach der ersten Erhebungsphase statt. Grund für ein zweites Interview mit der

Programmleitung war, dass sich diese einerseits dafür bereit erklärte und andererseits hierdurch gezielt weitere Informationen bzgl. der Programmtheorierekonstruktion und den spezifischen Aufgabenfeldern und individuellen Informationen der Programmleitung in Erfahrung gebracht werden konnten. Dieser Leitfaden wurde in drei thematische Bündel unterteilt: Beginnend mit Fragen zur Kooperation von ENB mit der Schule, anschließend mit programmspezifischen Fragen und abschließend mit Fragen zu GmS als Finanzgeber und übergeordnete Organisation. Allerdings wurde dieser Leitfaden gezielt offen und möglichst wenig strukturiert gestaltet, um eher ein interaktiveres Gespräch zwischen der Programmleitung und dem Autor dieser Studie zuzulassen. Hierdurch wurde es dem Autor ermöglicht, individuelle Schwerpunkte während des Interviews zu setzen, die spontan während des Gesprächs aufkamen. Hintergrund war, dass durch den anhaltenden Evaluationsprozess der Autor bereits weitere inhaltliche Recherchen getätigt hatte und bereits die Interpretationen aus der ersten Interviewphase mit in das Gespräch nahm. Somit konnten spontan auftretende Themen gezielt vor diesem Hintergrund um vertiefende oder ergänzende Fragen bereichert werden. Für einen genaueren Überblick wird auf den Leitfaden im Anhang verwiesen.

5.3.5 Datenaufbereitung

Nach Durchführung der Interviews mussten die Daten zunächst transkribiert und anschließend kodiert werden. Hierfür wurden zwei verschiedene Softwares genutzt: Zur Transkription der aufgenommenen Interviews wurde die Software "aTrain" verwendet, welche von der Universität Graz entwickelt wurde (siehe Haberl et al., 2024). Die dadurch gewonnene Transkripte wurden durch den Autor dieser Studie mit den Audioaufnahmen der Interviews abgeglichen und wo nötig korrigiert. Hierbei wurde sich an dem Transkriptionssystem von Dresing und Pehl (2018) orientiert, welche den Fokus auf die semantisch-inhaltliche Ebene legen und sprachliche Merkmale außen vor lassen. Darüber hinaus wurde zur Generierung der Kategorien im Rahmen der qualitativen Inhaltsanalyse nach Mayring dessen Open-Source-Software *QCAmap* verwendet (Mayring & Fenzl, 2024).

5.4 Ergebnisse

Die während der qualitativen Inhaltsanalyse deduktiv und induktiv gewonnenen Kategorien werden im Folgenden dargestellt. Hierbei werden zu jeder Kategorie Erläuterungen und zum Teil Zitate angeführt. Allerdings soll es sich hierbei lediglich um die Ergebnisdarstellung handeln. Es werden die im Rahmen der qualitativen Inhaltsanalyse beschriebenen Schlussfolgerungen zugelassen, allerdings geschieht die Interpretation dieser Ergebnisse vor dem Hintergrund der Fragestellungen in Kapitel 6.

Aufgrund dessen, dass die Interviewtranskripte nicht veröffentlicht werden und um Anonymität für die Interviewten zu bewahren, wird ebenfalls auf anonymisierte Personenreferenzen verzichtet. Dies hat den Grund, dass sich die Fachkräfte in – wie Zeitraum, in dem für das Projekt gearbeitet wurde, fachliche Spezialisierung und daraus folgender inhaltlicher Orientierung – wodurch durch Kenntnis dieser Eigenschaften die dahinterstehende konkrete Person identifiziert werden könnte. Gleichzeitig werden lediglich jene Zitate dargestellt, die nach Einschätzung des Autors dieser Studie keinerlei Rückbezüge auf dahinterstehende Personen zulassen.

Die im Rahmen der qualitativen Inhaltsanalyse entwickelten Kategorien wurden zur besseren Strukturierung in vier Oberkategorien eingeordnet:

1. Das Programm ENB
2. Der Programmkontext
3. Das Personal des Programms
4. Evidenzbasierung des Programms

Insbesondere der vierte Punkt bezieht sich auf das in 5.3.4 erläuterte Schülerprofil (siehe Anhang) und das in 5.3.1 dargestellte CIT-Interview. Die vier Oberkategorien wurden jeweils in verschiedene Unterkategorien aufgeteilt.

Zur Exemplifizierung und Explizierung der jeweiligen Kategorien werden Zitate aus den Interviews dargestellt. Diese werden immer in Anführungszeichen gesetzt und kursiv dargestellt. Teilweise werden einzelne Wörter, Wortgruppen oder Sätze in den Fließtext integriert, gleichzeitig werden ganze Sätze oder Passagen – die als besonders hervorhebenswert erscheinen – in Abgrenzung zum Fließtext dargestellt. Unabhängig davon wird bei der Explizierung der Kategorien und der Darstellung der Aussagen des Fachpersonals stilistisch der Konjunktiv verwendet und/oder die Sachverhalte aus der Wahrnehmung des Fachpersonals beschrieben. Dies soll die Aussagen des Fachpersonals nicht relativieren, sondern lediglich sprachlich ausdrücken, dass es sich hierbei um eine Meinung des Fachpersonals handelt.

5.4.1 Programm

Programm- & Angebotsgestaltung:

Im Rahmen der Interviews wurden zentrale Elemente des Programms und der Angebote angesprochen. Das Fachpersonal äußerte einhellig, dass es die große Freiheit, die ihm vom Programm in Bezug auf die Angebotsgestaltung gewährt wird, als sehr positiv und als Anerkennung der eigenen (professionellen) Kompetenz wahrnimmt: *„Das wird eigentlich komplett in meine Kompetenz gegeben, was eigentlich schön ist, was auch ein Vertrauen ist“*. Darüber hinaus wird von einem Teil des Fachpersonals die im Vergleich zur Klasse kleine Gruppengröße in den Förderstunden als sehr positiv wahrgenommen. Hierbei wurde u. a. geäußert, dass die Umsetzung des eigenen professionellen Anspruchs und des damit verbundenen Qualitätsverständnisses mit kleineren Gruppen besser realisiert werden könne: *„Also es ist immer so, je kleiner die Gruppe, desto besser“*. Als Begründung der reduzierten Gruppengröße wird der soziale Hintergrund der Kinder angeführt: *„Aber da war es so speziell, dass einfach durch die ganzen sozialen Themen, die in der Klasse sind, dass die großen Gruppen [...] einfach schon viel zu groß waren“*. Im Rahmen der Interviews wurde ebenfalls das Qualitätsmanagement angesprochen. Einerseits wurde es als wichtig erachtet, für zukünftige Programme das durch das Fachpersonal erworbene Erfahrungswissen zu sichern, zum Beispiel mittels Dokumentationen oder auch Ansprechpersonen für zukünftige Programmbeteiligte. Das Fachpersonal sieht sich dabei allerdings ressourcentechnisch (Zeit, Kompetenzen) nicht hinreichend ausgestattet, um das Qualitätsmanagement selbstständig durchzuführen. Am geeignetsten für diese Stelle aus dem Personalpool wird die Programmleitung gesehen, da sie den besten „Überblick“ über den Kurs hat und gleichzeitig aufgrund ihres engen Austausches mit den anderen Mitarbeitenden als Multiplikator dienen kann. Andererseits wurde darauf Bezug genommen, dass Stellen des Qualitätsmanagements

in GmS abgebaut wurden, wodurch die „*Expertise fehlt*“, dies seitens ENB umzusetzen. Das im Rahmen dieser Stelle durchgeführte Qualitätsmanagement wurde darüber hinaus als Ermöglichung der Umsetzung der Programmkonzeption beschrieben: „*Die engeren Kriterien von Qualitätsmanagement waren nicht immer erfüllt. Aber es ging halt schon darum zu schauen, wie kann man innerhalb des vorgegebenen Rahmens bleiben. Wie kann man die einzelnen Konzeptbausteine sinnvoll umsetzen. Ohne irgendwas zu vergessen und zu übersehen.*“

Neben den konkreten Programmbestandteilen wurde ebenfalls gefragt, wie die Programmangebote konkret vorbereitet werden. Das Fachpersonal beantwortete dies größtenteils folgendermaßen: Zur typischen Unterrichtsvorbereitung wird zunächst passendes Material gesucht, das zur Verfügung steht und dessen Nutzung im Unterricht vorstellbar erscheint. Anschließend werden methodisch-didaktische Überlegungen daran angeschlossen, auf welche Weise das gefundene Material in den Unterricht eingebaut werden könne. Die Überlegungen des Materials, aber auch die Antizipation der methodisch-didaktischen Umsetzung orientiert sich stark an der Wahrnehmung der Bedürfnisse der Schülerinnen und Schüler und der Identifikation von dazu passenden Förderinhalten. In diesem Zuge wurde beschrieben, dass deren Bedürfnisse und daraus abgeleitete Lerninhalte hinsichtlich ihrer Lebenswelt betrachtet werden, wobei die Bedürfnisse aufgrund der langen und positiven Beziehung mit den Schülerinnen und Schülern bekannt sind. Weiterhin wurde erwähnt, dass darauf geachtet wird, möglichst keine Monotonie bei der Auswahl des Materials entstehen zu lassen und darüber hinaus dieses methodisch möglichst abwechslungsreich in den Unterricht zu integrieren. Die Ideen für die methodische und inhaltliche Gestaltung des Unterrichts stammen nach Aussagen des Fachpersonals zum Teil aus Erfahrungen aus anderen Arbeitsstellen, an der die jeweilige Person parallel oder in der Vergangenheit beschäftigt wurde. Gleichzeitig wird sich dabei stark auf die eigene professionelle Kompetenz bezogen, die methodische und inhaltliche Abwägungen leitet. Die Konzeption der Unterrichtseinheiten wird als Ausdruck und Anwendung „eigener Kreativität“ beschrieben, die sich aus diesem Zusammenspiel der eigenen Erfahrungen, professionellen Kompetenzen und den jeweiligen Bedürfnissen der Kinder ergibt: *“Ich habe mich tatsächlich da sehr von den Themen der Kinder leiten lassen, von ihren sozialen Themen auch.”*

Teilweise wurde auch angesprochen, dass insbesondere darauf geachtet wird, den Unterricht spielerisch zu gestalten sowie erfahrendes und erlebendes Lernen zu ermöglichen, wobei möglichst positive Emotionen gefördert werden sollen. Gleichzeitig wurde hierbei angemerkt, dass es teilweise zu Abwägungsprozessen kommt zwischen der Korrektur von grammatikalisch oder lexikalisch inkorrekten Aussagen der Kinder und deren emotionalem Erleben und dem dadurch entstehenden positiven Bezug zum Lerngegenstand. In diesem Kontext wurde auch die Bedeutung von kindgerechter Sprache und Kommunikation hervorgehoben.

Programmziele & Angebotsqualität:

Des Weiteren sollen nun jene Inhalte dargestellt werden, die sich auf die Ziele von ENB und der damit verbundenen Angebotsqualität beziehen. Dies bezieht sich rein auf die Aussagen, die während des Interviews getroffen wurden und soll lediglich diese Perspektive darstellen. Da die Programmziele und die Anforderungen an die Angebotsqualität primär durch die Programmverantwortlichen definiert werden, handelt es sich somit um eine deskriptive

Beschreibung der geäußerten Zieldimensionen. Allerdings lassen sich hierdurch ggf. Argumentationsmuster und Sichtweisen des Fachpersonals herausarbeiten. Im Rahmen der Ziele wurde sich explizit auf die Zielgruppe von ENB bezogen. Hierbei wurden verschiedene Eigenschaften genannt, die die Zielgruppe spezifizieren sollten: Die Zielgruppe wird vom Übergang der Grundschule in die weiterführende Schule definiert: *„Dann haben wir neun Kinder, die die Noten nicht haben, weder fürs Gymnasium noch der Realschule. [...] Das sind die, mit denen wir am intensivsten gearbeitet haben. Das ist schon unsere Hauptklientel“*. Durch diese zukünftigen Bildungsperspektiven, die als negativ und als selbstwertschädigend wahrgenommen werden, wird eine Demotivierung der Kinder hinsichtlich ihres eigenen Bildungsverlaufes wahrgenommen. Da Kinder nicht nach diesem Kriterium im Grundschulalter selektiert werden können, wird in diesem Zuge die Eigenschaft der sozialen Benachteiligung angeführt, die mit diesem Übergang größtenteils assoziiert wird:

„Im Fokus, wenn es um Bildungsgerechtigkeit geht, sind für uns die Kinder, die von zu Hause her nicht die Unterstützung haben. Die, die es jetzt aufs Gymnasium geschafft haben, kriegen zu teilen, schon auch von zu Hause mehr Unterstützung, als der Personenkreis der Kinder, die da nicht dazu gehören.“

Soziale Benachteiligung wird allerdings teils unterschiedlich definiert, aber mehrheitlich wird sich auf das Vorhandensein eines Migrationshintergrundes und das Fehlen der elterlichen Unterstützung bei Bildungsangelegenheiten verwiesen, woraus sich somit die Zielgruppe und der damit verbundene Förderbedarf begründet:

„Wir sind jetzt hier in einem Stadtteil mit einem extrem hohen Migrationsanteil. Das heißt, wir haben jetzt in der Klasse auch, ich sage jetzt mal im Grunde genommen, 19 Kinder, haben mindestens ein Elternteil, der einen nichtdeutschen Hintergrund hat.“

„Aber letztlich, ich versuche einmal in der Woche mit jedem von diesen Kindern, die einen besonderen Förderbedarf haben, so zu arbeiten, wie es eigentlich zu Hause zwei, dreimal in der Woche passieren müsste.“

Konkret wird somit die Auswahl der Zielgruppe und auch der Förderanspruch des Programms mit dem bereits in Kapitel 2.3 theoretisch erläuterten Begriff der Bildungsgerechtigkeit verknüpft. Insbesondere soll das Programm diejenigen Kinder fördern, die die benötigte Förderung nicht an anderer Stelle erhalten, wobei hier die Schule und die Eltern genannt werden. Der Schule wird insbesondere eine mangelnde Vermittlung der für die weitere Schulbildung benötigten Grundfertigkeiten attestiert:

„Das Hauptaugenmerk ist für uns, merke ich schon, und auch für mich die Gruppe der Kinder, die über die Jahre immer mehr abgehängt wird. Die dann an den Mittelschulen aufschlagen und mit den einfachen Rechenarten nicht zurechtkommen. Die schlecht oder ungenügend lesen. Die keine ganzen Sätze formulieren.“

Aufgrund der hohen Homogenität der Klasse hinsichtlich ihrer wahrgenommenen sozialen Benachteiligung wird unter Bildungsgerechtigkeit insbesondere die Förderung der leistungsschwächeren Kinder beschrieben. Leistungsstärkere Kinder, die als sozial benachteiligt wahrgenommen werden und sich innerhalb des Klassenverbandes befinden, sollen *„implizit“* durch die Teilnahme am Programm gefördert werden: *„Wir haben auch implizit die gefördert, die aufs Gymnasium gehen, durch die Sprachförderung. Da haben wir über die Jahre schon viel gemacht.“*

Im Rahmen der Interviews wurde selbstkritisch darauf verwiesen, dass das Ziel, die Eltern der geförderten Kinder durch ENB ebenfalls in das Programm zu integrieren, nicht auf die Weise erreicht werden konnte, wie angestrebt. Insbesondere wurde hier wahrgenommen, dass die Eltern unzureichend über das Programm und seine Ziele informiert wurden und gleichzeitig wenig Initiative zeigten, mit den Programmverantwortlichen ins Gespräch zu kommen. Die Programmbeteiligten schlussfolgerten daraus, dass für zukünftige Programme die Elternarbeit stärker strukturell in das Programm aufgenommen und an der Schule institutionalisiert werden sollte. Gleichzeitig wurde überlegt, inwiefern mit den Eltern eine Vereinbarung getroffen werden könne, sodass mindestens einmal im Schuljahr ein Gespräch zwischen den Programmbeteiligten und den Eltern stattfinden könne. Im Laufe der Interviews ergaben sich hinsichtlich des Umgangs mit den wahrgenommenen Förderbedarfen, die sich durch die schulischen Mängel ergeben, zwei Strategien: Einerseits sollen schulische Mängel durch das Programmangebot kompensiert werden und andererseits soll möglichst produktiv auf das „System Schule“ eingewirkt werden, um es u. a. zum Umgang mit sozialen Benachteiligungen zu befähigen. Insbesondere die Ergänzung der schulischen Mängel werden durch das Fachpersonal elaboriert: Einhellig wird auf die Wahrung oder Herstellung der sozio-emotionalen Lernvoraussetzungen hingewiesen, die durch die Schule nicht oder in unzureichender Form verwirklicht würde. In diesem Zusammenhang wird oftmals darauf verwiesen, dass die Anzahl der Schülerinnen und Schüler einer Klasse zu groß sei und die Zeit der Lehrkraft zu gering, um deren sozio-emotionalen Lernvoraussetzungen zu wahren oder aber auch nur zu diagnostizieren; als notwendige Voraussetzungen für die Diagnose und Durchführung von Maßnahmen hinsichtlich der sozio-emotionalen Lernvoraussetzungen wird die Beziehung zu dem Schüler oder der Schülerin genannt. Als weiteres Argument wird angeführt, dass das Programm eine „Auszeit“ für die Kinder schafft, in der zwar etwas gelernt wird, aber nicht auf schulische Art: *„So diese ganzen Sachen sind so ganz kleine Puzzleteile, wo du einfach merkst, okay, die kommen da gerne hin, das ist für sie eine schöne Auszeit, sie lernen was. Und können das dann immer auf ihren Schulalltag und auch auf die anderen Angebote, die wir hier machen, anwenden.“* Mit „Auszeit“ wird sich insbesondere auf das Nicht-Vorhandensein von Leistungsdruck und Bewertungsnormen bezogen, denen die Schülerinnen und Schüler in der Wahrnehmung des Fachpersonals im Schulalltag ausgesetzt seien. Dazu wurde ebenfalls der Bezug zum „spielerischen Lernen“ gezogen, bei dem es nicht um Leistung, sondern um Spaß ginge, *„und trotzdem spielerisch [zu] lernen, ohne es wirklich zu merken und nicht so dieses Frontallernen“*. In diesem Kontext wird das schulische Lernen als „Frontallernen“ wahrgenommen. Dies bezieht sich teilweise auch auf die schulischen Inhalte, die als sehr „technisch“ wahrgenommen werden und wenig „Kreativität“ fördern.

All diese Punkte beziehen sich (implizit) auf Ziele des Programms, indem schulische Mängel definiert und teils Antworten darauf formuliert werden. So wird erwähnt, dass ein „Raum ohne Bewertungen geschaffen“ werden solle und ein kreativer und ansprechender „Zugang zu den schulischen Inhalten“ eröffnet werden kann. Aufgrund des fehlenden Zwangs zur Bewertung wird es möglich, den Kindern Anerkennung zuteilwerden zu lassen: *„Du musst jedem Kind, unabhängig von seinen Anlagen, von seinen Voraussetzungen und was von zu Hause her kommt, immer wieder das Gefühl geben, dass auch sein Beitrag wichtig ist.“* Darüber hinaus wird vereinzelt die Förderung von Talenten als wichtiges Programmziel gesehen, wobei der hier verwendete Talentbegriff einen Kontrast zum wahrgenommenen schulischen Talentbegriff formuliert:

„Wir haben so viele Kinder mit großen Talenten. Die schulisch jetzt nicht die Überflieger sind, die dann bei Dreiern und Vierern landen. Aber wo mein Gefühl ist, die hat so viele andere

Talente und so gute Eigenschaften, was immer das ist. Frustrationstoleranz, Ausdauer und so weiter und so fort. Das wird bestimmt was aus seinem Leben machen. Ich glaube, es ist wichtig, dass wir die Kinder nach der vierten Klasse entlassen und dass kein Kind das Gefühl hat, jetzt ist es vorbei, jetzt habe ich nur noch die Mittelschule geschafft und mein Leben ist gelaufen und aus mir kann definitiv nichts mehr werden.“

Allerdings wird einer der Schwerpunkte des Programms auch in der Unterstützung der Schule in ihrem Bildungsauftrag gesehen: *„Förderung [richtet sich] natürlich nach den schulischen Erfordernissen. Das heißt, die Förderung läuft entlang des Lehrplans, entlang der schulischen Anforderungen, das heißt schwerpunktmäßig entlang der Wochenplanung.“* Das bedeutet, ein wichtiges Programmziel – welches auch strukturell in ENB verankert ist – wird als die Anpassung an die *„schulischen Erfordernisse“* und die Konzeption von dahingehenden Unterstützungsangeboten wahrgenommen. In der Wahrnehmung des Fachpersonals beinhaltet dies ebenfalls die Unterstützung und Förderung der Kinder hinsichtlich der Schulfächer: *„Aber wir haben auch natürlich den Anspruch, sie bei den Leistungen in Deutsch, in Mathe usw. zu verbessern oder ihnen zu helfen, dass sie sich verbessern oder ihnen zu helfen, dass sie den Anschluss nicht verlieren. Das ist ein ganz wesentliches Erfolgskriterium.“* Durch die Unterstützung des Unterrichts der Lehrkraft und die dadurch verbundene Teilnahme an dessen Unterricht wird darauf Bezug genommen, dass dadurch für das Programm wertvolle Perspektiven auf die Kinder gewonnen werden können, die letztlich in die pädagogischen Entscheidungen und den Förderalltag integriert werden können. Wichtig erscheint es hierbei auch, für die Kinder *„sichtbar“* zu sein:

„sondern indem ich auch einfach mit im Unterricht bin, wenn Neues durchgenommen wird, wie am Mittwoch zum Beispiel, dass ich dann einfach da bin und wenn es darum geht, das im Anschluss zu üben und zu vertiefen oder wie auch immer, dass wir das dann auch gemeinsam haben, zu zweit, zu dritt und dass das dann einfach auch für die Kinder sichtbar wird.“

Die Unterstützung der Schule bei ihrem Bildungsauftrag soll u. a. durch die kontinuierliche Zusammenarbeit mit den schulischen Verantwortlichen erfolgen. Hierbei wird allerdings erwähnt, dass dieses Zieles noch nicht vollständig erreicht wurde: *„Trotzdem wäre es das Endziel bzw. die höchste Form dieser Zusammenarbeit, wenn wir es schaffen würden, die Lehrerin zu einem Teammitglied zu machen. Das heißt, dann wären unsere regelmäßigen Absprachen zu viert oder zu fünft, je nachdem wie viele Leute da mit im Boot sind, der Idealzustand.“* Eine mögliche Zusammenarbeit mit dem schulischen Personal, insbesondere mit der Lehrkraft, wird auf verschiedenen Ebenen verortet: Grundsätzlich wird der kommunikative Austausch mit der Lehrkraft erwähnt, der meistens über informelle „Tür-und-Angel-Gespräche“ erfolgt. Zur Begründung wurde angeführt, dass aufgrund der unterschiedlichen Stundenpläne und einem fehlenden formalen Rahmen keine formalen Absprachen bzgl. fester Termine zum kommunikativen Austausch realisiert wurden. Innerhalb dieses informellen Austausches wurden allerdings verschiedene Themen angesprochen: Bei der Planung der Förderzeiten für ein Schuljahr mussten insbesondere mit der Lehrkraft konkrete Zeiten und Tage festgelegt werden, an denen das Programm ihr Angebot in der Schule offerieren konnte. Neben jenen allgemeinen organisatorischen Absprachen wurden insbesondere Absprachen mit der Lehrkraft hervorgehoben, die das pädagogische Vorgehen des Programms bestimmten: So wurden einerseits im gegenseitigen Austausch mit der Lehrkraft die Kinder für die Einzelförderung bestimmt sowie bei Feststellung temporärer Lernschwierigkeiten eine erhöhte Teilnahme an der Einzelförderung festgelegt. Diese Zusammenarbeit wird allerdings nicht nur auf die Lehrkraft bezogen, sondern auf alle

schulisch relevanten Akteurinnen und Akteure: „Auch das ist ein Erfolgskriterium oder ein Qualitätsmerkmal, dass man einfach mit allen Akteuren in einem engen Austausch ist.“ Hierunter fallen u. a. – wie bereits erwähnt – die Eltern der Kinder.

Das besondere Profil des Programms wird insbesondere in Hinblick auf die vorhandenen sozialen Beziehungen zwischen dem Fachpersonal und den Kindern skizziert. Diese werden häufig mit der Beziehung des Fachpersonals zu den Kindern beschrieben und dem Programmpersonal wird dabei eine andere, meist längerfristige und intensivere Beziehungsqualität – kontrastierend zur Lehrkraft – diagnostiziert: „Also Bildungsqualität [...] steht und fällt mit Beziehung. Es geht um gewachsene, es geht um gute Beziehungen zwischen Erwachsenen und Kind, es geht um professionelle Beziehungen natürlich, um eine gesunde Nähe und Distanz.“ Eine Begründung für die zentrale Rolle der Beziehung wird als durch sie entstehende „Beziehungsmotivation“ beschrieben: „Kinder lernen und erbringen Leistung, einfach auch aus den Beziehungsmotivationen heraus zum Erwachsenen. Und wenn das eine gute, eine gesunde, eine funktionierende Beziehung ist, zu einem Kind oder umgekehrt zu einem Erwachsenen, dann motiviert das ganz anders.“ Die Begründung der Rolle von Beziehungen in Bezug auf schulischen Erfolg wurde häufig auch mit eigenen Erfahrungen beispielhaft belegt:

„dass eine erwachsene Person, die an dich geglaubt hat und dich irgendwie positiv beeinflusst hat, es ausmachen kann, wie dein Leben sich entwickelt und wie sich auch deine Lernlaufbahn entwickelt und so weiter und das denke ich ist einfach wichtig, dass sie das merken und dass da jemand da ist und klar, im Idealfall sollten das die Eltern machen, im Idealfall sollten es die Lehrer machen, aber beide haben viele andere Themen auch noch und andere Kinder oft auch in den Familien und deswegen ist es wichtig, dass es so etwas gibt, wo das eben genauer darauf geachtet wird, denke ich.“

Ziel des Programms ist es somit auch, dass das Fachpersonal den Kindern als Bezugspersonen und Vorbildern dienen. Insbesondere die durch solche Beziehungen entstehende „Freude“ wird als ein weiteres Motiv verstanden, soziale Beziehungen mit den Kindern aufzubauen:

„Freude am schulischen Alltag. Viele Dinge, die sie bei uns machen und tun, da sehen wir schon, dass sie die mit Freude erfüllen. Ich meine, Freude macht es auch, wenn ich mit einem Erwachsenen, mit der Lehrkraft in einem Flow bin, in einer guten Beziehung bin. Alles, was Freude bereitet, sollte auch Maßstab an der Schule sein. In der Schule geht es immer so trocken [zu].“

Zum Beziehungsaufbau und zur optimalen Förderung der Kinder wird häufig idealtypisch auf die Einzelförderung verwiesen. Diese wird als „1-1-Situationen“ bezeichnet, in welchen individuell auf die (sozio-emotionalen) Lernvoraussetzungen der Kinder eingegangen und in denen gleichzeitig eine „Atmosphäre des Vertrauens“ geschaffen werden kann. Diese 1-1-Situationen werden insbesondere als beziehungsbegünstigend beschrieben: „Also da ist es einfach für Kinder, die nicht mit der deutschen Sprache aufwachsen, einfach nochmal wichtiger, dass wir irgendwie 1-1-Situationen schaffen. Das geht hier an der Schule tatsächlich nur über so individuelle Förderstunden oder individuelle Unterstützungsstunden, so muss man es eigentlich nennen.“ In diesem Sinne wird größtenteils durch das Fachpersonal geäußert, dass die optimale Betreuungsform die individuelle Förderung sei und dass sich durch die Reduzierung der Gruppengröße daran angenähert würde. Allerdings werden diesen 1-1-Situationen eine besondere Qualität zugesprochen und insbesondere wird in

diesem Kontext erneut Bezug auf den herkunftsbedingten Hintergrund der Kinder genommen. Durch diese individuelle Zuwendung seitens von ENB können auch diejenigen Kinder Zuwendung erfahren, die im alltäglichen schulischen Unterrichtsgeschehen diese eher weniger erfahren: *„Ich glaube, dass wirklich kein Kind hinten runterfällt, dass alle Kinder im Blick sind, dass über jedes der Kinder, die gerade in Not oder in Schwierigkeiten sind, auch gesprochen wird, dass sich auch ausgetauscht wird.“*

Neben der Beziehungsqualität wird insbesondere die Motivation der Kinder als zentraler Einflussfaktor auf Lernprozesse wahrgenommen. Wie bereits erwähnt, wurde das Konzept Motivation mit dem Konzept der Beziehung kombiniert und als „Beziehungsmotivation“ beschrieben. Darüber hinaus wird ebenfalls Bezug auf das Alter der Kinder genommen: *„Und in der Grundschule sind sie noch so jung, da kannst du sie noch gut motivieren. Und das sehe ich bei diesen Kindern hier sehr und das macht mir viel Hoffnung für die weiter.“*

5.4.2 Kontext

Rahmenbedingungen:

Im Laufe des Interviews wurden ebenfalls inhaltliche Bezüge zu den Rahmenbedingungen und Kontextfaktoren hergestellt, in denen das Programm verortet ist. Im Folgenden werden diese dargestellt. Einheitlich wurde sich vom Fachpersonal auf die herkunftsbedingte und familiäre Situation und daraus abgeleitete Förderbedarfe bezogen. Ein Argumentationsmuster war die Beschreibung der Mängel der kindlichen Fähigkeiten aufgrund bestimmter herkunftsbedingter Faktoren:

„Und viele von diesen Kindern, mit denen wir hier oder allgemein bei GmS arbeiten, haben viele von diesen Basic Skills nicht von daheim bekommen, aus was für Gründen auch immer, es ist vielleicht kulturell anders, als es in Deutschland ist, die Eltern haben selber sehr viele Probleme, die Eltern arbeiten wahnsinnig viel, es sind vielleicht nicht alle Elternteile vorhanden, da gibt es ja verschiedene Gründe.“

Das bedeutet, dass konkrete wahrgenommene Mängel in der Förderung seitens des Elternhaushaltes diagnostiziert werden, worauf das Programm seinen Fokus richten sollte: *„Und ich finde, das muss immer mehr im Fokus genommen werden, weil die Bedingungen, weil die Bildungs- und Erziehungsarbeit oft in den Elternhäusern weniger geleistet wird.“*

Des Weiteren wurde Bezug auf die Corona-Pandemie genommen, die den Programmverlauf aus Wahrnehmung des Fachpersonals entscheidend beeinflusst hat. Hierbei wurde argumentiert, dass durch die Corona-Pandemie die Zeit im Elternhaushalt verlängert wurde und somit schulische Förderangebote – insbesondere in jenen Haushalten, in denen wenig schulbezogene Förderaktivitäten stattfinden – nicht realisiert werden konnten.

Integration in Schule:

Im Folgenden sollen nun jene Aussagen dargestellt werden, die die Wahrnehmung des Personals hinsichtlich der Integriertheit von ENB in der Schule beschreiben. Integration wird hier verstanden als (wahrgenommene) strukturelle Einbettung, Anerkennung und Interaktionsqualität und -häufigkeit zwischen den Programmbeteiligten und den schulischen

Akteurinnen und Akteuren. Zuvor soll allerdings dargestellt werden, welches Verständnis die Programmbeteiligten von Schule als Institution haben, da angenommen werden kann, dass dieses Verständnis einerseits Ausdruck der Integriertheit in die Schule ist und andererseits die Art der Wahrnehmung der tatsächlichen Integriertheit beeinflussen könnte. Viele Programmbeteiligte definieren ihre eigene Rolle im Programm und im Bildungswesen komparativ zu den typischen schulischen Akteurinnen und Akteuren: Den Lehrkräften. So wird teilweise versucht, sich methodisch und bildungstheoretisch von diesen abzugrenzen. In diesem Kontext wird oftmals ein schulisches Bildungsqualitätsverständnis bzw. einzelne Bestandteile dessen herangezogen, um das eigene Bildungsqualitätsverständnis zu kontrastieren. Größtenteils wird hierbei der Aspekt des Bewertens genannt, wobei insbesondere das Anwenden einer einheitlichen Bewertungsnorm auf alle Schülerinnen und Schüler ohne Berücksichtigung derer individueller Lernvoraussetzungen und sozialer und familiärer Lage kritisch betrachtet wird. Teilweise werden auch schulische Diagnose- und Selektionsverfahren stark mit eigenen Idealen und Qualitätsverständnissen gegenübergestellt:

“Oder sage ich, gute Bildung ist das, wo Kinder oder überhaupt Menschen etwas für sich persönlich mitnehmen und manchmal, und das ist meistens so, ist ein wirklicher Wissensgewinn oder ein Erkenntnisgewinn, nicht unbedingt sofort abrufbar [ist] durch Testschreibung. Und Bildung, meiner Meinung nach, wird immer noch so begriffen, was kann ich praktisch über Wissensabruf eruieren.”

In diesem Zusammenhang wurde des Öfteren das schulische Qualitätsverständnis als primär auf Wissensvermittlung ausgerichtet beschrieben. Eigene Bildungsqualitätsverständnisse orientierten sich hingegen eher an den Konzepten der Persönlichkeitsentwicklung und des Kompetenzerwerbs. Die Lehrkraft wird demnach teilweise als Teil des Systems wahrgenommen, die in dieser vom System ihr zugeteilten Rolle gefangen sei und diese Rolle auch stets mit sich „transportiere“. Durch das Programm und dessen Integration in die Schule würden somit neue professionelle Rollen und Akteurinnen und Akteure in das Schulsystem aufgenommen, die somit auch das „System Schule verändern können“.

Vor diesem Hintergrund wird die Einbettung des Programms in die Schule als weniger ausgeprägt wahrgenommen. Insbesondere die seitens der Schule offerierten Rahmenbedingungen werden teilweise als unzureichend betrachtet: So wird die Zeit, die für die einzelnen Angebote zur Verfügung gestellt wird, größtenteils als zu kurz bemessen – vornehmlich vor dem Hintergrund der Teilung der Klasse in kleinere Gruppen, was dazu führt, dass die zur Verfügung stehende Zeit pro Gruppe reduziert wird. Hierbei lässt sich ein Konflikt zwischen den beiden Aspekten *Zeit* und *Individualität* auffinden, die sich abhängig der gegebenen Rahmenbedingungen konträr gegenüberstehen. In anderen Worten: Will man die Zeit für die Förderung der einzelnen Kinder steigern, so müsste man diese in größere Gruppen oder im Klassenverbund unterrichten. Andererseits wird hierdurch der Anspruch der individuellen Förderung unterminiert, der sich durch die Reduzierung der Gruppengröße auszeichnet, wobei sich dadurch die Förderzeit für jedes Kind entsprechend der Gruppenanzahl reduziert. Darüber hinaus wird teilweise angemerkt, dass die räumlichen Möglichkeiten eingeschränkt sind und die Räume nicht nach eigenen Vorstellungen gestaltet werden können. Allerdings wird in diesem Kontext auch erwähnt, dass die Schule bestimmte Abwägungsprozesse realisieren muss, da sie ebenfalls ihre Ziele hinsichtlich des Lernplans umsetzen möchte. Das bedeutet, dass es teilweise als Abwägungsprozess verstanden wird, ob bei Ressourcenfragen stärker die Ziele der Schule oder des Programms berücksichtigt

werden sollten. In diesem Kontext wird die Schule als „komplexe Organisation“ beschrieben, wobei es an Zeit und Aushandlungsprozessen bedarf, um Innovationen und neue Strukturen zu integrieren:

„Aber das ist, ich sage immer, Schulen sind schwierige Partner. Nicht weil sie per se schwierig sind oder weil die anderen Personen, Akteure irgendwie schwierig sind als Menschen oder so. Aber es ist halt ein sehr komplexes Gebilde und da arbeiten viele Menschen und es sind viele Kinder, die betreut und denen gerecht werden muss. Und von daher, also da gerät schnell mal was aus dem Blick, was sozusagen noch Randgeschäft ist, sage ich mal. Und ja, das ist durchaus eine Herausforderung, finde ich.“

Durch den Begriff „Randgeschäft“ wird die oben dargestellte Abwägungsfrage hinsichtlich der Berücksichtigung der Ziele der Schule oder des Programms adressiert, wobei klar dargestellt werden soll, dass die Rolle von ENB an der Schule als „Randgeschäft“ wahrgenommen wird. Diese Wahrnehmung spiegelt sich vor allem in der Aushandlung von Räumen, Zeitslots und Ressourcen zwischen dem Programm und der Schule wider. Das bedeutet, es wird konkret auf die Aushandlungsprozesse zwischen dem Programm und der Schule verwiesen. Das Programm müsse insbesondere seine „anderen [methodischen] Zugänge“ mit der Schule aushandeln und „Überzeugungsarbeit“ leisten. Dazu müssten die schulischen Verantwortlichen hinsichtlich des Programms, dessen Methoden und Ziele „sensibilisiert“ werden. Allerdings wird der Schule auch ein Lernprozess unterstellt, da diese nach Aussage der Programmbeteiligten bisher keine Erfahrungen mit externen Kooperationspartnern hätte und die Schule sich so auf die neue Situation „eingewöhnen“ müsste und man zunächst „zueinander finden“ müsste.

„Wenn wir uns als Team regelmäßig treffen und uns über die Kinder austauschen, über den aktuellen Lernstand, über Termine usw. Aber da sind wir weitergekommen. Aber es braucht immer ein bisschen Zeit. Man kann nicht davon ausgehen, dass man sich das am Schuljahr anfangs so hinstellt und dann läuft es auch jetzt. Das muss man auch zueinander finden.“

Aus dieser Wahrnehmung heraus wird Bezug auf den Umgang mit der Klassenlehrkraft genommen: Der Umgang seitens ENB wird als sehr bedacht beschrieben und als sehr orientiert auf den Bedürfnissen der Lehrkraft und den schulischen Beteiligten. Es wurde des Weiteren ausgesagt, dass aufgrund der mangelnden Erfahrung der Schule mit externen Kooperationspartnern mit der Lehrkraft gemeinsame Formen des Umgangs miteinander und Rollendefinitionen gefunden werden müssten. Ein großes Anliegen des Fachpersonals war es, möglichst die Lehrkraft bei der Unterrichtsgestaltung zu entlasten.

5.4.3 Personal

Einstellung gegenüber dem Programm:

Das Personal und die Lehrkraft wurden ebenfalls nach ihrer Einstellung über das Programm befragt. Hierbei wurden verschiedene Aspekte erwähnt, allerdings wurde einhellig eine sehr positive Einstellung geäußert. In diesem Kontext wurden oftmals auch Vergleiche zu anderen, ähnlichen Programmen oder anderen schulbezogenen Erfahrungen gezogen, um die positive Einstellung zu begründen. Es wurden größtenteils übereinstimmend Eigenschaften des Programms genannt, die als besondere „Stärken“ des Programms wahrgenommen werden: Hierunter fallen insbesondere der fehlende Leistungsdruck, die Vermittlung von Freude am Lernen, die positive und langfristige Beziehung mit den Kindern und die Ermöglichung einer

„Auszeit“ vom schulischen Lernen. Insbesondere auch die Arbeit im Team wird als sehr positiv, herzlich und wertschätzend wahrgenommen.

Das Konzept des Programms wird als kohärent beschrieben, insofern als die einzelnen „Programmbestandteile“ so miteinander verzahnt seien, dass diese „Sinn machen“:

„Wir haben ein klares Tableau, wir haben drei Gruppentermine, wir haben drüber hinaus die Differenzierungs- und Einzelförderstunden. Das macht Sinn. Wir merken auch selber, dass das sich jetzt gut eingefügt hat in den Stunden. Das macht Sinn, was wir machen, das hat Hand und Fuß.“

Einheitlich wird auch der Umfang des Programms und die von den Finanzgebern zur Verfügung gestellten Ressourcen als überdurchschnittlich wahrgenommen. Implizit wurde hierbei auch ein bestimmtes Qualitätsniveau geäußert, das aufgrund der überdurchschnittlichen Ressourcenausstattung mehr leisten kann als andere Programme oder die Schule als Institution: *„Ja, die 60.000 Kosten sämtlich. Wenn man sich das mal vor Augen führt. Nur in einer Klasse.“*

Wie bereits angeklungen werden mit dem Programm auch eigene Erwartungen hinsichtlich eines Bildungsqualitätsverständnisses und der Institution Schule verbunden. Dem Programm wird in diesem Zusammenhang „Mut“ attestiert, andere und neue Wege bei der Bildung der Kinder zu gehen. In diesem Kontext werden Programme im Allgemeinen als zentrale Akteurinnen zur Schulentwicklung beschrieben und nicht mehr Schulen. Programme seien diejenigen, die diesen „Mut“ aufbrächten, kontrastierend zu den schulischen Akteurinnen und Akteuren. Dieser „Mut“ wird insbesondere im Zusammenhang mit den Inhalten und Methoden des Programms assoziiert, die nicht primär in Schule vorkommen und die mittels des Programms strukturell in die Schule verankert würden. Diesbezügliche Veränderungswünsche hinsichtlich des Schulsystems werden teilweise lediglich an Stiftungen und Programme adressiert; unter Auslassung der Schulen.

Eigene Rollenbeschreibung:

Neben den Wahrnehmungen des Programms und deren Rahmenbedingungen wurden ebenfalls die Wahrnehmungen des Fachpersonals der eigenen (professionellen) Rolle befragt. Ein Teil des Fachpersonals beschrieb einen Prozess des „Sich-Findens“, der beim Beschäftigungseintritt in das Programm auftritt und sich auf die eigene Rollenbeschreibung und die daraus abgeleitete didaktisch-methodische Unterrichtsgestaltung bezieht. Dies begründe sich in der großen „Freiheit“, die das Programm dem Fachpersonal hinsichtlich der Angebotsgestaltung zuteilwerden lässt. Diese Freiheit wird zwar, wie beschrieben, als äußerst positiv wahrgenommen, bringe allerdings diesen Findungsprozess mit sich: *„Im ersten Jahr war ich echt ein bisschen am Verzweifeln. Ich habe gedacht, was mache ich eigentlich?“* Hierbei wurde erwähnt, dass sich zunächst auf die Rahmenbedingungen angepasst und ein passendes Format gefunden werden müsse. Die eigene Rollenbeschreibung wird dabei oftmals von der eines typischen Lehrers abgegrenzt. Die Rolle von Lehrkräften wird als stark reglementiert wahrgenommen; im Gegensatz zur eigenen „Freiheit“.

Die eigene professionelle Kompetenz, die die Rolle bestimmt, wird einhellig durch die eigenen Erfahrungen in verschiedenen schulischen und außerschulischen Anwendungsgebieten begründet. Neben den Erfahrungen wird zum Teil auch die eigene

professionelle Ausbildung angeführt, die stark das eigene Rollenverständnis prägt. Insbesondere die dort erworbenen Methoden werden stark mit der eigenen Rolle im Programm assoziiert. Von einem Teil des Fachpersonals wurde wiederum angemerkt, dass die vom Programm vorgegebenen Rahmenbedingungen einheitlich mit dem eigenen Rollen- und Qualitätsverständnis konvergieren: *„und das hat einfach auch ganz gut zum Programm GmS gepasst, dass über Bindung, über Raumgestaltung, Ortgestaltung, wie man mit den Kindern ins Gespräch kommt, auch mit Ritualen, also das sehr gut zu dem Thema GmS gepasst hat.“*

Diagnostische Urteile & Wirksamkeitsurteile:

Im Rahmen dieser Arbeit sollte untersucht werden, inwiefern das Fachpersonal diagnostische Urteile in ihrer Praxis beschreibt und inwiefern damit Wirksamkeitsurteile formuliert werden. Darüber hinaus wurden vom Fachpersonal Aussagen über die (personenbezogenen) Bedingungen der diagnostischen Urteile getätigt. Diagnostische Urteile werden in dieser Arbeit als Aussagen über epistemologische Prozesse, die Eigenschaftszuschreibungen der Kinder erlauben, definiert. Wirksamkeitsurteile andererseits beziehen sich auf die Produkte erwarteter Kausalzusammenhänge zwischen dem pädagogischen Intervenieren und (diagnostizierten) kindlichen Eigenschaften. Wirksamkeit ist in diesem Zusammenhang positiv konnotiert und benötigt dabei immer den Bezug auf explizite oder implizite Qualitätsverständnisse des Fachpersonals.

Das Fachpersonal äußert verschiedene, aber methodisch ähnliche Quellen für ihr diagnostisches Wissen. Als primäre Quelle werden Beobachtungen in konkreten Praxissituationen beschrieben. Allerdings wird der Fokus auf jeweils unterschiedliche Aspekte des kindlichen Verhaltens gelegt: In diesem Sinne werden einerseits sprachliche Äußerungen der Kinder herangezogen, um diagnostische Urteile zu fällen. Hierbei werden Konzepte wie sprachliche Eloquenz, emotionale Befindlichkeit oder soziale Konflikte interpretativ direkt oder indirekt aus den sprachlichen Äußerungen der Kinder diagnostiziert. Auf nicht-sprachlicher Verhaltensebene werden häufig Persönlichkeitseigenschaften, Intelligenz, Motivation, Begabung, Auffälligkeiten im Verhalten, Unterrichtsstörungen, Aufgabenbearbeitungen sowie Freude an Aktivitäten wahrgenommen und daraus diagnostische Urteile gebildet.

„Wir haben schon einen kleinen Kreis an 4 bis 5 Kindern. Du setzt dich hin, du erklärst es, du machst es gemeinsam. Es funktioniert irgendwie, schriftliches Subtrahieren. Es klappt dann plötzlich. Und dann macht das Kind noch eine Aufgabe, noch eine Aufgabe. Und freut sich, es hat es verstanden. Oder du machst schriftliches Dividieren. Und du erklärst es immer wieder und machst es immer wieder. Nach einer Stunde klappt es dann plötzlich. Das Kind freut sich. Und drei Tage später setzt du dich hin mit dem Kind. Und es ist wie ausgelöscht. Du merkst schon, es bleibt nicht bei jedem Kind mit der Intensität hängen. Manche Kinder haben einfach andere kognitive Voraussetzungen.“

Die Beobachtungen werden größtenteils in Situationen beschrieben, in denen die Programmbeteiligten in direkter Interaktion mit den Kindern stehen. In Hinsicht auf die Diagnostik des Vorhandenseins einer (positiven) sozialen Beziehung mit den Kindern wurde zum Teil darauf verwiesen, dass dies erkannt werde, wenn sich *„die Kinder einem öffnen und mit Fragen kommen“*. Darüber hinaus wird der Austausch der Lehrkraft als weitere Informationsquelle beschrieben, worauf sich eigene diagnostische Urteile stützen. So gibt die

Lehrkraft insbesondere über die Aufgabenbearbeitung im Rahmen des Wochenplans Auskunft.

Allerdings wird von einem Teil des Fachpersonals auch die Präferenz geäußert, keinerlei oder nur ausgewählte Informationen über die Schülerinnen und Schüler durch andere Personen mitgeteilt zu bekommen:

„Zu viele Vorabinformationen, finde ich, neigt man dann dazu, Kinder dann nicht mehr so offen zu betrachten. Und schon voreingenommen und auch mit gewissen unbewussten Erwartungen oder eben Nicht-Erwartungen an sie heranzugehen. Während meine Erfahrung ist, je mehr ich über ein Kind überhaupt nichts weiß und ich ganz offen an es herangehe, ist es manchmal erstaunlich oder toll für die Kinder, einfach total neutral genommen zu werden und ihnen begegnen zu können.“

Informationen von außen werden in diesem Sinne als urteilsbeeinflussend beschrieben, insofern als das Urteil in eine bestimmte Richtung gelenkt wird: *„Also ich habe sehr wenig Informationen über die Schülerinnen. Und das ist auch okay. Also ich brauche es nicht unbedingt. Bei manchen wäre es dann ganz gut.“* Falls Bedarf an zusätzlichen Informationen über die Schülerinnen und Schüler bestünde, so würde man sich an eine geeignete Ansprechperson richten, die über jene Informationen verfüge: *“wo man mal nachfragen kann, oder wenn gerade ganz prekäre familiäre Situationen sind, warum ein Kind zum Beispiel sehr auffällig wird oder so.“* Allerdings wird in diesem Zusammenhang der Wunsch geäußert, dass beim Kennenlernen der Kinder zunächst keine weiteren eigenschaftsbeschreibenden Informationen vermittelt werden sollten, um die Kinder unvoreingenommen kennenlernen zu können. Denn ansonsten könne es dazu führen, dass die Kinder *„stigmatisiert“* werden würden. Ausnahmen werden bei Leserechtschreibschwäche, Autismus, Phobien, spastische Beeinträchtigung oder ähnlichem beschrieben. Dabei wird geäußert, dass *“wenn ein Verhalten auffällig ist, [dann ist es] manchmal ganz gut, eine Information zu haben [die erlaubt, das] Verhalten besser verstehen und einordnen“* zu können.

Darüber hinaus wurde durch einen Teil des Fachpersonals die Idee geäußert, dass die diagnostischen Urteile des Fachpersonals dokumentiert werden könnten, um den diagnostischen Prozess zu strukturieren und gleichzeitig einer Programmevaluation zugänglich zu machen. In Verbindung mit geäußerten diagnostischen Urteilen wurden häufig auch individuelle und subjektive Bezugsnormen angewandt, um das Leistungsverhalten der Kinder einordnen zu können. Diese stammen aus eigenen Erwartungen, die sich aus Erfahrung mit dem Kind gebildet haben, sowie Erfahrungen aus anderen Programmen aus der Vergangenheit und Gegenwart (auch aus höheren Jahrgangsstufen). Der Autor dieser Arbeit postuliert in diesem Zuge, dass diagnostische Urteile darüber hinaus nötig sind, um Wirksamkeit zu erkennen. In den Aussagen der Fachkräfte lassen sich Argumentationsmuster erkennen, die hinreichende und notwendige Bedingungen für (programmbezogene) Wirksamkeitsurteile skizzieren.

Hinreichende Bedingungen für Wirksamkeitsurteile:

Als hinreichende Bedingungen wird u. a. die Lernzeit der Kinder beschrieben. Allein die Anwesenheit der Kinder in den Förderstunden und die dadurch geschaffene Lernzeit wird als *„ein großer Mehrwert für die Kinder“* aufgefasst. Hierbei wird nicht expliziert, welche zeitliche, inhaltliche und qualitative Dimensionen darüber hinaus gegeben sein müssen, sondern lediglich beschrieben, dass die bloße Lernzeit bereits als in diesem Sinne wirksam

beschrieben wird. Ebenfalls wird die Zuwendung zu den Kindern als hinreichende Bedingung aufgefasst, in diesem Sinne wirksam zu sein: *„Aber ich glaube, je mehr Leute da sind, die die Kinder sehen und diese für die nächste Stufe mitnehmen, und sei es nur durch "Ich seh dich!" ist total wertvoll. Und dadurch passieren einfach viele andere Dinge. Und ich glaube schon, dass die Kinder in der Klasse mehr Zugang bekommen haben zu schulischen Themen. Auch wenn es schwer messbar ist.“*

Notwendige Bedingungen für Wirksamkeitsurteile:

Als notwendige Bedingungen für Wirksamkeitsurteile werden verschiedene Dinge beschrieben:

- Die Übereinstimmung der im Unterricht behandelten Inhalte mit den kindlichen Interessen.
- Positive und über einen längeren Zeitraum gewachsene Beziehungen.
- Sozio-emotionale Stabilität der Kinder, verstanden als Fehlen von lernhinderlichen sozio-emotionalen Ausprägungen.

Diese drei Bedingungen werden allerdings unabhängig voneinander genannt, aber beinahe einheitlich von dem Fachpersonal als notwendig für (personenbezogene) Wirksamkeitsdimensionen wahrgenommen. In diesem Kontext wurden konkrete Beobachtungen von dem Fachpersonal angeführt, um die Wirksamkeit ihrer pädagogischen Handlungen zu diagnostizieren: Auf sozio-emotionaler Ebene werden insbesondere jene Beobachtungen genannt, nach denen die Kinder ihr Verhalten selbst korrigieren, bestimmte Begriffe in sprachlichen Äußerungen verwenden, Spaß an den Aktivitäten zeigen oder aber motivierter wirken. Das bedeutet, diese Beobachtungen werden angeführt, um das Vorhandensein bestimmter (impliziter) Wirksamkeitsdimensionen zu beschreiben: *„Also die Kinder erzählen mir ja auch viel darüber, wie toll sie das finden, wie gern sie da hingehen und was sie dann machen und so weiter. Das ist, glaube ich, schon toll und allgemein würde ich sagen, ist es ein Erfolg, weil diese Kinder deutlich motivierter wirken, als sie es anfangs waren.“*

Gleichzeitig werden schriftliche Produkte der Kinder hinsichtlich dieser und weiterer (nicht geäußertes) individueller Kriterien und Erwartungen verglichen und bei Übereinstimmung oder Übertreffen ein positives Wirksamkeitsurteil gefällt wird.

Zusammenfassend lässt sich somit festhalten, dass diagnostische Urteile und Wirksamkeitsurteile vornehmlich auf den eigenen oder kommunizierten Wahrnehmungen des Fachpersonals beruhen, die in der Praxis gewonnen wurden. Andere Kriterien, wie Noten oder Diagnostiktests, werden hierfür nicht angeführt. Auf Nachfrage wird einerseits die Aussagekraft von Noten stark in Zweifel gezogen und die Güte der eigenen Wahrnehmung wird dabei kontrastiert. Der objektive Wirksamkeitsnachweis mit (standardisierten) Diagnostiktests wird in diesem Sinne als „schwierig“ beschrieben: *„Also in dem Sinne, genau vergleichen kann man es nicht. Weil, selbst wenn man eine andere Klasse nimmt, die hat ja auch wieder andere Voraussetzungen. Also man wird es ja nie richtig beweisen können, in Anführungszeichen. Aber ja, ich finde, man merkt es.“*

5.4.4 Evidenzbasierung

Im Folgenden werden jene Ergebnisse dargestellt, die im Rahmen der CIT-Interviews gewonnen wurden (siehe Kapitel 5.3.1 sowie 5.3.4). Daran vorangestellt werden unter *Umgang mit quantitativen Diagnostiktests* lediglich jene Aussagen, die den breiteren Kontext erläutern sollen. Anschließend werden jene Kategorien dargestellt, die in der qualitativen Inhaltsanalyse konstruiert wurden.

Umgang mit quantitativen Diagnostiktests:

Ein Anliegen dieser Studie ist es, einen Beitrag zur evidenzbasierten Entwicklung des Ganztages zu leisten. Hierbei wurde zunächst die Erfahrung und Wahrnehmung der Programmbeteiligten hinsichtlich des Umgangs mit den quantitativen Diagnostiktests (ELFE, SELLMO) befragt. Diese sind – wie in Kapitel 3 beschrieben – integraler Bestandteil des Programmkonzepts. An die quantitativen Diagnostiktests werden verschiedene Erwartungen gestellt. Einerseits werden sie eingesetzt, um das Programm zugänglich für wissenschaftliche Begleitungen zu machen und dem Programm nach außen einen gewissen „*wissenschaftlichen Anstrich*“ zu geben. In diesem Kontext wird die wissenschaftliche Begleitung als „*Steckenpferd*“ des Programms beschrieben. Andererseits sollen diese Wirksamkeitsnachweise hinsichtlich des Programms zulassen: „*Es ist wichtiger als alles andere, dass wir in irgendeiner Form Nachweise liefern, dass das Programm wirkt.*“

Allerdings wird die Erhebung der quantitativen Diagnostiktests teilweise auch als Belastung für die durchführenden Akteurinnen und Akteure aufgefasst. Dies bezieht sich zunächst auf den konkreten Ressourceneinsatz, um die Daten zu erheben und auszuwerten. Dies kostet Zeit, Personal und findet gleichzeitig zum Teil in den Förderstunden statt, wodurch die bereits als kurz wahrgenommene Zeit am Kind noch weiter verkürzt wird. Insbesondere wurde hierbei erwähnt, dass die Tests teilweise mehr Zeit in Anspruch nähmen, da einige Kinder sprachliche Verständnisschwierigkeiten bei dem Ausfüllen der Items hätten, wodurch die Programmbeteiligten unterstützend zur Seite stehen müssten. Gleichermaßen wurde geäußert, dass partiell reflektiert wird, inwiefern die jeweiligen situativen Rahmenbedingungen Einfluss auf das Testergebnis haben könnten, vor dem Hintergrund, dass die Testergebnisse zur Bewertung des Programms herangezogen werden. So werde auch darüber nachgedacht, inwiefern die jeweilige Situation möglichst optimal gestaltet werden kann, sodass die Testergebnisse nicht beeinträchtigt würden: „*Und dann will man schon irgendwie für diese Testung dann auch optimale Bedingungen generieren, damit die Kinder auch wirklich ihre beste Leistung abliefern.*“ Dies lässt sich damit begründen, dass die Gestaltung der Testsituation und die Auswahl des Zeitpunktes durch das Fachpersonal entschieden werden kann, wodurch jene Entscheidungsräume entstehen. Allerdings wird die Bedeutung der Diagnostiktests auf Programm- und Geldgeberebene teilweise relativiert, da Interpretations- und Bewertungsprozesse vornehmlich über mündlichen, „*zwischenmenschlichen*“ Austausch erfolgt und die Ergebnisse der Diagnostiktests entsprechend argumentativ kontextualisiert und somit begründet werden können.

Verständnisfragen:

Nach Vorlage und Erläuterung des Schülerprofils (siehe Anhang) wurden unterschiedliche Typen von Aussagen festgestellt. So wurden unter anderem Verständnisfragen bzgl. des

Schülerprofils gestellt. Die Inhalte dieser Fragen könnten Aufschluss darüber geben, welche Informationen noch zur Verfügung gestellt werden müsste, um das Verständnis des Schülerprofils zu erhöhen:

Es schien keine systematischen Verständnislücken durch die Erläuterung und die Interpretation des Schülerprofils über alle Personen hinweg zu geben, sondern jede Person stellte individuelle Verständnisfragen, die bestimmte Inhalte tangierten: Es wurde nach weiterführenden Informationen zu der Vergleichsgruppe gefragt, um die prozentualen Werte der Lernmotivation und des Wortverständnisses besser einordnen zu können. Auch wurde mehrfach nachgefragt, mit welcher Methode die Lernmotivation operationalisiert und erhoben wurde. Ein weiterer Typ von Verständnisfragen war die kommunikative Absicherung des eigenen Verständnisses des Schülerprofils. So wurden Begriffe und Interpretationen aus der dargebotenen Erläuterung nochmals wiederholt und expliziert sowie anschließend nachgefragt, ob das eigene Verständnis korrekt sei. Diese Art von Verständnisfragen bezogen sich vor allem auf die Interpretation der Lernkurve, die abgebildeten Noten und zur Klärung der verlangten Interpretation des Schülerprofils.

Aussagen über quantitative Testinstrumente:

In Bezug auf das Testinstrument wurden verschiedene Aussagen getätigt, die bestimmte Annahmen explizierten oder Einstellungen offenbarten.

Messbarkeit von Konstrukten:

In Zusammenhang mit dem Schülerprofil und in Bezug auf quantitative Testinstrumente im Allgemeinen wurde mehrmalig auf die Messbarkeit der dahinterstehenden Konstrukte referiert. So wird die Einstellung gegenüber bestimmten Diagnoseinstrumenten zur Erhebung der Schülereigenschaften oder deren Leistungsverhalten abhängig von Messbarkeitsüberlegungen gemacht:

“Also ich muss sagen, mit diesen anderen Teilen kenne ich mich jetzt zu wenig aus. Also ich habe schon mal gehört, was da läuft und so weiter und so fort. Also ich glaube [...] bei den Lesekompetenzen, glaube ich, sieht man auch wirklich gute Erfolge. Was können sie lesen, wie ist das Leseverständnis und so weiter. Also das, glaube ich, kann man daran sehr gut messen.”

Insbesondere hinsichtlich der Lesekompetenz wurden beinahe einheitlich quantitative Diagnoseinstrumente als geeignet betrachtet, um Erfolge messbar zu machen. So wurde den Diagnostiktests im Allgemeinen eine gewisse „Messbarkeit“ unterstellt, wobei dieses Urteil stark von den Zielkonstrukten und den individuellen Erfahrungen bzgl. der Diagnostiktests abhängig ist: *„man kann natürlich einen gewissen Standard schon erheben, im Vergleich zu anderen Schulen. Und der Test bleibt ja gleich, also am Ende sieht man ja dann schon einen Fortschritt. Also ich glaube schon, dass man mit solchen Tests gewisse Fortschritte messen kann.“*

Gleichzeitig wurde noch eine gewisse Skepsis oder Vorsicht bzgl. der Operationalisierung und Messung von Lesemotivationen geäußert, da hier wenig Erfahrungswerte vorliegen würden. Darüber hinaus werden Lesekompetenzen als „verschulter“ wahrgenommen:

“Also ich denke, dass das natürlich bei, also kann ich mir vorstellen, da kenne ich mich nicht genug aus, aber dass das bei so Sprach- und Leseförderungssachen einfacher ist, weil das vom Naturell her verschulter ist wahrscheinlich einfach.” Schule wird hier stärker mit messbaren Konstrukten assoziiert.

Spezifizierung der Aussagekraft der Diagnostiktests hinsichtlich Schulnoten:

In diesem Kontext wird die Aussagekraft der Diagnostiktests auch mit den schulischen Noten verglichen: *“Ja, aber ich wüsste jetzt keine Alternative. Also, weil die eigenen Schulnoten, die bringen jetzt weniger, würde ich behaupten. Und ich meine, solche Tests sind ja erstellt worden mit allen Kompetenzen in [den] einzelnen Noten [...]. Die Proben beziehen sich ja meistens nur auf Teilkompetenzen, was [im Unterricht fokussiert wurde]. Und dieser Test fragt ja alles ab, auch beim Lesen. Lesegeschwindigkeit, Leseflüssigkeit, Lese-Sinnentnahme und so weiter. Deswegen finde ich schon, das kann man heranziehen.”* Hierbei wird ein eher holistisches Testverständnis geäußert, nach dem Lesekompetenz als Ganzes gemessen und welches mit den Schulnoten kontrastiert wird, die lediglich auf „Teilkompetenzen“ abzielen.

Vergleich mit eigenem diagnostischen Urteil:

Neben diesen Überlegungen hinsichtlich der Messbarkeit und des Vergleichs mit den Schulnoten wurde allerdings eingewandt, dass der Informationsgehalt der Diagnostiktestergebnisse oftmals als redundant wahrgenommen würde, da man aufgrund eigener diagnostischer Beobachtungen und Urteile bereits über diese Informationen verfüge:

“Die Erfahrung zeigt, dass die Tests eigentlich nicht wirklich etwas anderes aussagen als das, was man täglich im Unterricht erlebt. Meistens spiegelt es das wider. Es gab einzelne Kinder, wo wir tatsächlich erstaunt waren. Da denke ich gerade an einen Fall, wo ich dachte, ach, das ist eigentlich ganz gut. Und der Test war viel schlechter als das, was ich eingestuft hätte. Es gibt Einzelfälle, wo es zusätzlich ein Informationsgehalt gibt, wo man genauer hinschaut.”

Die Aussagekraft bzw. der Informationsgehalt der Diagnostiktests wird hierbei relativiert und lediglich für Einzelfälle als sinnvoll erachtet: *“Ja, in Einzelfällen. In Summe sind diese Tendenzen eigentlich schon immer deckungsgleich mit dem, was man so erlebt.”*

Allerdings wird der allgemeine Nutzen der Diagnostiktests hinsichtlich des eigenen diagnostischen Urteils stark angezweifelt bzw. es wird eine gewisse Unentschlossenheit ausgedrückt, inwiefern dieser als sinnvoll erachtet werden sollte. Vielmehr wird für das eigene diagnostische Urteil argumentiert, welches sich aus den Erfahrungen in der Praxis und der Interaktion mit den Kindern ergibt: *“Also ich bin so zwiegespalten, weil wie gesagt, ich erlebe die [Kinder] jede Woche, das heißt, ich sehe auch so was passiert. Das zeigt mir es nochmal schwarz auf weiß. Das wäre sicher nicht verkehrt. Ich weiß es nicht, ob es am Ende des Tages so Mehrwert bringt.”* Somit wird insbesondere das eigene „Erleben“ der Kinder als wichtiger diagnostischer Zugang erachtet und dem Informationsgehalt der Diagnostiktests gegenübergestellt.

Diagnostiktests als Unterstützung des eigenen diagnostischen Urteils:

Von einem Teil des Fachpersonals wird allerdings ebenfalls eingeräumt, dass jene Diagnostiktests sich dafür eignen, eigene Mängel im diagnostischen Urteil zu identifizieren und zu korrigieren und mögliche Förderbedarfe abzuleiten:

“Es ist oft eine Stütze, weil ja man kennt die Kinder, aber man hat ja auch nicht selber alles im Kopf von allen einzelnen Schülern. Und gerade so Motivationstests habe ich jetzt vorher auch noch nie verwendet. Also das finde ich jetzt für mich auch ganz interessant. Ich habe diesen Test nie gesehen, ich würde es interessieren, wie das erhoben wird. Klar kann es hilfreich sein, wenn man so etwas auf einen Blick sieht, wie die Kinder sich entwickelt haben. Auch um zu entscheiden, welche Schülerinnen noch gefördert werden können, oder auch wie man sie fördern kann.”

Zweifel an Aussagekraft der Diagnostiktests:

Neben der Relativierung der Aussagekraft der Diagnostiktestergebnisse werden ebenfalls konkrete Zweifel an deren Aussagekraft geäußert. So wird zum Teil die Vergleichbarkeit der durch ENB geförderten Klasse mit der Normstichprobe bezweifelt: *“Ja, ich meine gut, die Vergleichsgruppen sind natürlich dann, der ELFE-Test ist ja erstellt mit Vergleichsgruppen von ganz Deutschland. Und dann ist es natürlich wieder schwierig, weil ich finde, es müssen schon noch mehr Faktoren mit einbezogen werden, also was soziale Herkunft und so weiter angeht.”* Darüber hinaus werden auch konkret Zweifel geäußert, wenn die Testergebnisse nicht den eigenen Erwartungen entsprachen:

“Also zum Beispiel das, dass das Absenken auf die dritte Klasse unter dieses Niveau der ersten Klasse und hier oben das zur Zweiten war, das würde mich so extrem misstrauisch machen. Ich kann mir nicht vorstellen, dass Menschen Worte verlieren oder Sprachverständnis verlieren. Also unter gewissen Umständen vielleicht lange im Ausland gewesen und nicht mehr diese Sprache sprechen, vielleicht als aktiven Wortschatz oder bei einem Unfall, aber das verstehe ich nicht. Das glaube ich nicht. Also ich glaube, dass das Sprachverständnis durchaus weitergeht und sich erweitert.”

Allerdings wurde in diesem Zitat nicht der Vergleich zur Normstichprobe herangezogen, sondern die intraindividuellen Vergleiche mit den Vorjahren, was darauf schließen ließe, dass es zu einem Missverständnis hinsichtlich des Vergleichs mit der Normstichprobe gab und die Werte als absolute Leistungswerte interpretiert wurden.

Zweifel an Sinnhaftigkeit der Items der Diagnostiktests:

Des Weiteren wird die Güte der im Test verwendeten Items und der dahinterstehenden Konstrukte von einem Teil des Fachpersonals kritisch betrachtet:

“Ich finde das ziemlich, mir tut das total leid, also ein Kind auf 32,6 Lernmotivation. Auch das würde ich sehr mit, wenn die sagt, machst du gerne Hausaufgaben? Nö, ich spiele natürlich lieber draußen oder mache irgendwas, wenn ich nach Hause komme, als dass ich diese Hausaufgaben mache. Ich sehe ja auch, was sie für Hausaufgaben haben. Schematisches Einfüllen von irgendwelchen Zeug ist super langweilig, blöd. Weiß auch nicht, ob es wirklich was hilft. Ich meine, Üben ist manchmal schon wichtig, für einige Kinder sogar wichtiger als für andere, je nachdem, wo ihre Stärken oder Schwächen sind.”

Hierbei wird die Relevanz und Aussagekraft des Items bzgl. der Hausaufgabenbearbeitung kritisch betrachtet.

Interpretation:

Das Fachpersonal wandte unterschiedliche Strategien an, um die Informationen aus dem Schülerprofil in den Programmkontext zu setzen und dahingehend zu interpretieren.

Identifikation von Ursachen:

Dabei wurde mehrheitlich zunächst nach Ursachen gesucht, um den dargestellten Abfall der Schülerin des Schülerprofils im Wortverständnis im Vergleich zur Normstichprobe zu erklären. Dieser Abfall bzw. der allgemeine Verlauf der Lernkurve wird mehrheitlich als ungewöhnlich oder überraschend bezeichnet. Eine mögliche Ursachenquelle wurde zum großen Teil im familiären Umfeld situiert:

“Also ich denke, wenn sich das so massiv verändert und auch noch verschlechtert, obwohl die Einzelförderung da ist, dann würde ich jetzt mal, weil ich ja viel von dem Projekt halte, nicht daran denken, dass das am Projekt liegt, sondern würde eben gucken, ist da irgendwas passiert. Ich weiß ja, dass es bei ein paar Schülerinnen massive Einschnitte gab, privater Natur, und damit hat es oft zu tun. Also die klassischen Sachen wie Scheidung oder einfach Gewalt in der Familie oder andere Konflikte, teilweise auch Tod oder es gibt ja ganz schlimme Dinge, die junge Menschen schon erleben müssen. Damit hat es oft zu tun, dass natürlich dann, wie ich vorher auch schon gesagt habe, Lernen als allerletztes hinten ansteht und die Sachen auch nicht mehr gelernt werden.”

Gleichzeitig wird hierbei Bezug auf die Klassenkomposition genommen: *“Ich würde erstmal schon auch schauen, wo ist der Mensch dahinter? Waren da Themen? Ich habe jetzt ein Kind im Gedanken, auf das das zutreffen könnte. [...] Ich würde da sofort spiegeln, wer könnte das sein? Es sind halt oftmals sehr familiäre Themen. Wir haben bei 18 Kindern 18 Schicksale, muss man so sagen. Das ist hier sehr extrem, wenn man so will.”*

Die allgemeine familiäre Situation der Kinder der Klasse wird als allgemein „sehr extrem“ bezeichnet und könnte in der Wahrnehmung des Fachpersonals eine plausible Ursache für den Abfall im relativen Wortverständnis sein.

Schrittweises Interpretieren:

Der Interpretationsprozess an sich gestaltet sich teilweise sehr strukturiert und reflektiert, wobei teilweise „laut“ gedacht wird:

“Die Skala ist jetzt nicht so eindeutig für mich. Ich würde mir sagen, in der 2. Klasse war es gut und in der 3. ist es schlechter als 1. Klasse, würde ich vermuten. An dem Tag hat das Kind einfach null Bock gehabt. Ich würde es erst einmal verifizieren wollen, wer es ist und was wirklich dahintersteckt. Dass natürlich die Deutschnote auch noch abgesackt ist. Von 4 auf 5. Wie gesagt, es ist für mich eigentlich so ein Anzeichen, dass da in der Motivation einfach auch, wenn es jetzt gerade eben auch in der 3. Klasse, unter dem Klassendurchschnitt 30 %, das ist schon sehr wenig, das zeigt mir eigentlich, dass das Kind keinen Zugang zur Schule hat und zum Lernen. Da würde ich versuchen, einfach mal zu eruieren, woran liegt es wirklich und wie kann man es in der Einzelförderung bestärken?”

Im Verlaufe des Interpretationsprozesses werden unterschiedliche Ursachen genannt und miteinander abgewägt und letztlich wird mit einer Frage das weitere, mögliche Vorgehen beschrieben. Darüber hinaus werden weitere Informationen über die Person der Schülerin verlangt. Gleichzeitig wird hier der *“Zugang zu Schule und Lernen“* als Ursache angegeben

und auch die Einzelförderung als mögliche Maßnahme, dem zu begegnen. Es wurde von mehreren Programmteiligen geäußert, dass zur besseren Interpretation der Daten zusätzliche Informationen über die Schülerin benötigt würden, insbesondere über ihre familiäre Situation, welche häufig als mögliche Ursache für den Lernverlauf genannt wurde. In diesem Zuge wurde gleichfalls die Überlegung geäußert, dass jene psychologischen Tests, die die Schulpsychologin mit den Kindern durchführe, ebenfalls für die Interpretation herangezogen werden könnten, wobei hier sich explizit auf Konzepte wie allgemeine Intelligenz und Verarbeitungsgeschwindigkeit bezogen. Allerdings wurde dabei auch eingeräumt, dass man selbst nicht die Kompetenz hätte, dies adäquat zu interpretieren und in Relation zu setzen.

Schlussfolgernde Handlungsbegründung:

Das Fachpersonal ergänzte zum Teil seine Interpretation der Informationen des Schülerprofils mit Überlegungen, welche Maßnahmen sich daran anknüpfen ließen.

Allgemeines, unspezifisches Überlegen:

Dies geschah partiell dadurch, dass lediglich erwähnt wurde, dass geeignete Maßnahmen gesucht werden müssten, die die Lernmotivation und den Wortschatz fördern können. Daran anknüpfend wurde unspezifisch geäußert: *„ich würde schon die Lernmotivation irgendwie ein bisschen im Blick haben auch.“* In diesem Kontext wurde teilweise auch Bezug darauf genommen, dass man sich gemeinsam mit der Programmleitung bestimmte Maßnahmen überlegen würde – und diese somit in die eigenen Überlegungen miteinbeziehen würde.

Initiierung von Gesprächen mit der Schülerin:

Neben diesen allgemeinen und unspezifischen Überlegungen wurden größtenteils auch konkrete Überlegungen gegenübergestellt. Eine häufig vorkommende Schlussfolgerung war es, ein Gespräch mit der Schülerin zu initiieren und ihre Meinung zu belastenden Ereignissen zu fragen:

“zu gucken, okay, wie ist es auch in der Schule, wie empfindet sie das, die Kinder sind ja da oft sehr reflektiert, die wissen schon, dass es schlechter geworden ist oder dass sie das nicht mehr so gut verstehen und das stresst sie auch sehr. Und wenn es dann jemanden gibt, der mit ihnen darüber redet, dann ist das schon mal gut und dann kann man natürlich auch dementsprechend die Einzelförderung daran anpassen und alle anderen Sachen.“

Hieraus wird ersichtlich, dass zunächst ein diagnostischer Schritt in Erwägung gezogen wird, der Informationen über mögliche soziale oder familiäre Situationen generieren soll, die als eine mögliche Ursache für den beobachteten Leistungsabfall herangezogen werden könnte. Dies wird darin begründet, dass die Schülerinnen und Schüler allgemein in der Lage sind, ihren eigenen seelischen Zustand reflektieren und kommunizieren zu können. Hierbei wird ebenfalls die Einzelförderung als sinnvolle Maßnahme beschrieben. Dieses Schlussfolgerungsmuster wird häufiger vom Fachpersonal geäußert: *“Da würde ich versuchen, einfach mal zu eruieren, woran liegt es wirklich und wie kann man es in der Einzelförderung bestärken?“* Das bedeutet, als erster Schritt wird die Ursachenfindung erwogen, auf deren Grundlage anschließend die Einzelförderung geeignete Maßnahmen realisieren solle.

Reflexionsprozesse:

Anknüpfend oder vorausgelagert an die Interpretations- und Schlussfolgerungsüberlegungen ließen sich teilweise ad-hoc Reflexionen beobachten.

“Und deshalb ist es für mich auch so wichtig, je früher man die Kinder begleiten kann, dass diese Frustration und diese Motivation, die fehlende, gar nicht erst passiert. Weil das sehe ich schon in der Grundschule, dass viele Kinder in diese Frustration kommen und gar keine Lust mehr haben auf Lernen. Und wenn das mal passiert ist, ist natürlich das Wiederhochkommen schwieriger als je früher man anfängt. [...] Und deshalb denke ich mal, das wäre sicher zielführend für Stiftungen, auch schon früher einzusteigen, dass genau das passieren kann, dass solche Kinder schon früher in ihrer Lust auf Lernen und auf Sprache schon mal abgeholt werden, dass sie gar nicht erst in dieses Loch kommen und dieses Ich-hasse-Lernen. Das zeigt sich mir. Frühere Förderung, Einzelförderung und Schauen, das ist ja auch der Kernpunkt bei GMS, was ja auch das Schöne ist, dass immer auf das Kind geachtet wird, nicht so sehr, wie kriege ich den Schulstoff vermittelt, sondern was ist mit dem Kind, was gibt es für Hintergründe und wie kann ich es da abholen, wo es steht, um das andere wiederum zu fördern.”

Diese Reflexionsprozesse beziehen sich auf verschiedene Inhalte und teilweise wird das eigene Rollenverständnis oder aber das Programmverständnis reflektiert und ggf. präzisiert. Dabei kamen insbesondere eigene Erfahrungen und das eigene Rollenverständnis zum Tragen.

Aussagen über Störvariablen:

Die Interpretationen des Schülerprofils hinsichtlich der Aussagekraft der dort aufgeführten Testergebnisse wird zum Teil in Relation gebracht mit Überlegungen zu möglichen Störvariablen, die die Testergebnisse beeinflusst haben könnten.

Entwicklungs- und Beschulungseffekte:

So wurden die Ergebnisse hinsichtlich möglicher Entwicklungs- und Beschulungseffekte hinterfragt: *“Ich meine, ein gewisser Fortschritt ist ja irgendwie eigentlich auch gegeben, weil die Kinder werden ja älter, sie sind länger in der Schule, Programm-unabhängig oder nicht. Einen Fortschritt machen sie ja so oder so, finde ich. Ja, es ist natürlich dann schwierig eigentlich zurückzuführen, was hat das Programm dazu beigetragen.”*

Allerdings wurde eher Unklarheit dahingehend geäußert, wie aufgrund dieser möglichen Störvariable mit den Ergebnissen umgegangen und inwiefern durch die Ergebnisse Rückschlüsse auf das Programm gezogen werden können. Gleichzeitig lässt sich schlussfolgern, dass hier evtl. Verständnis hinsichtlich des Umgangs mit den in den Diagnostiktests verwendeten T-Werten bestehen.

Fehler in der Selbstbeobachtung der Schülerinnen und Schüler:

Als weitere mögliche Störvariable werden Fehler in der Selbstbeobachtung der Kinder beschrieben: *“weil es ist ja oft so [ist], dass die Schüler das dann nochmal anders erzählen oder auch anders empfinden oder selber auch nicht alles wissen und das irgendwie nur mehr so ein Gefühl ist, weiß man ja auch nicht.”* Den Schülerinnen und Schülern werden hierbei teilweise als (noch) nicht fähig zur Selbstbeobachtung beschrieben. Darüber hinaus wurden

Fremdbeobachtung als valider und zuverlässiger beschrieben als die Selbstbeobachtungen der Kinder:

„Ja, also ich denke mal, [...] also die Selbstbeobachtung ist oft viel kritischer [als die Fremdbeobachtung durch Fachpersonal], aber dass sie das wahrscheinlich auch so sehen würde, dass es ihr in der ersten und zweiten Klasse besser ging und dass es jetzt schwieriger ist, also das wissen die Kinder meistens schon ganz gut einzuschätzen.“

Hierbei wird dargestellt, dass einerseits die kindliche Selbstbeobachtung „*viel kritischer*“ sei als die Fremdbeobachtung durch das Fachpersonal, allerdings wird den Kindern auch die Fähigkeit zugesprochen, allgemein ihre Stimmung einschätzen zu können.

Tagesform:

Letztlich wird noch die Tagesform während der Testung als potenzielle Störvariable erwähnt: *„Von daher ist das natürlich auch, wenn das so eine Spitze hat, ist natürlich immer die Frage, welche Art von Wörtern sind das? Ist es auch eine Tagesform? Es gibt Kinder, die sich an einem Tag einfach komplett verweigern. Gerade bei den Extremen.“*

Praxistauglichkeit des Schülerprofils:

Im Rahmen der Interviews wurde zuletzt nach der Praxistauglichkeit des Schülerprofils oder ähnlichen Zusammenstellungen von Ergebnissen aus quantitativen Diagnostiktests gefragt. In diesem Zuge werden teilweise in den vorherigen Kapiteln bereits dargestellte Ergebnisse wiederholt und unter dem Punkt der Praxistauglichkeit zusammengeführt.

So werden *bereichsspezifisch* das Heranziehen quantitativer Testergebnisse als sinnvoll erachtet, wobei insbesondere Testergebnisse in Bezug auf die Lesekompetenz als geeignet betrachtet würden. Allerdings herrscht noch Skepsis vor hinsichtlich des Umgangs mit den Ergebnissen der Motivationstests. Hier wurden sich größtenteils konkretere Informationen bzgl. der Skalen und Items gewünscht. Des Weiteren sollten bei den Testungen zusätzlich *weitere Informationen* erhoben werden, um die Testergebnisse besser einordnen zu können. Hierbei wurden vor allem familiäre und soziale Gegebenheiten sowie die jeweilige Tagesform des Kindes angeführt. Ergänzend zu den Testergebnissen wurde darüber hinaus der Vorschlag unterbreitet, *Dokumentationen* von alltäglichen Beobachtungen der durch das Fachpersonal wahrgenommenen Lernfortschritte anzufertigen. Diese Informationen könnten dann entweder selbst zur Evaluation herangezogen werden oder die Interpretation der Testergebnisse ergänzen. Für die Auswertung und Interpretation der Test wurde teilweise angemerkt, dass dies aufgrund von mangelnden Zeitressourcen und Kompetenzen nicht selbstständig bewerkstelligt werden könne. Anstatt dessen wurde darauf verwiesen, dass hierbei auf eine *externe evaluierende Person* mit wissenschaftlichem Hintergrund zurückgegriffen werden sollte. Daran anknüpfend wurden die Durchführung und Erhebung der Diagnostiktests teilweise als *Belastung* beschrieben und als konkurrierend zu der „*Zeit am Kind*“:

„Ich weiß es nicht, ob es am Ende des Tages so Mehrwert bringt. Einfach, wenn man sieht, wo haben wir auch unsere Zeit, solche Auswertungen zu machen und sind wir dann vielleicht doch lieber direkt am Kind dran und merken es auch. Also ich bin da so zwiegespalten.“

Die Diagnostiktests werden auch teilweise als Möglichkeit gesehen, verschiedene *Perspektiven* auf das Kind zu bekommen und als Korrektiv der eigenen Perspektive zu dienen bzw. den Fokus auf Dinge zu richten, die evtl. nicht bedacht oder vergessen wurden.

6. Rekonstruktion der Programmtheorie

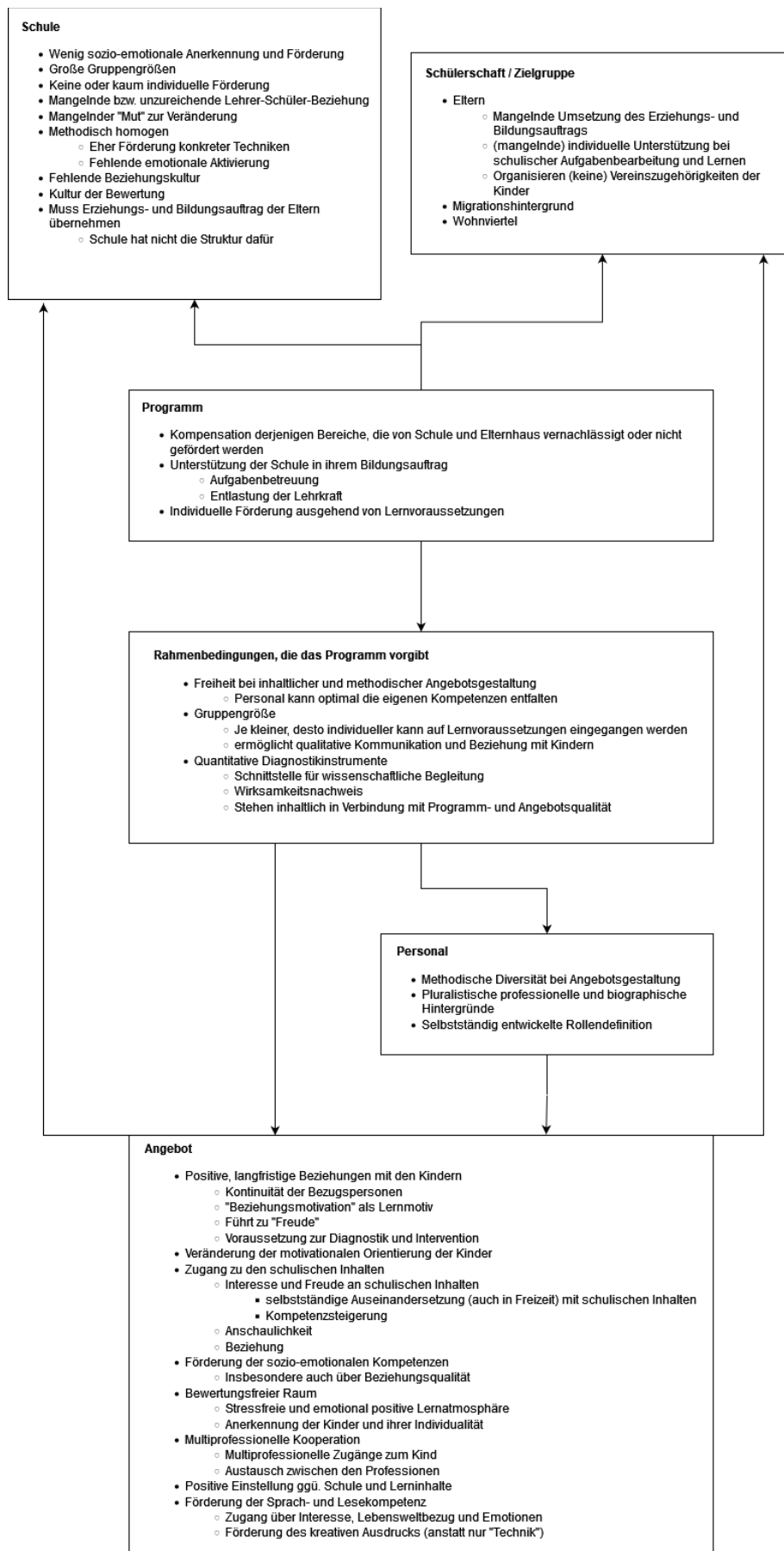
Im Folgenden wird die Programmtheorie dargestellt, die anhand der Aussagen, die in den Interviews mit dem Fachpersonal (unter Ausschluss der Lehrkraft) getätigt wurden. Da im Laufe der Interviews einige Aussagen getroffen wurden, die sich je nach Interpretation als programmtheoretische Annahmen beschreiben lassen, wurde stark selektiv vorgegangen. Hierbei wurde kein genau definiertes Selektionskriterium verwendet, sondern insbesondere darauf geachtet, dass die Annahmen von mindestens zwei interviewten Personen getroffen und dass diese von den Personen als „selbstverständlich“ dargestellt bzw. kommuniziert wurden. Dazu wurden insbesondere die Art der sprachlichen Darstellung, die Eindrücke aus den Interviewsituationen sowie die weiteren Erfahrungen, die der Autor dieser Arbeit im Austausch mit ENB gewonnen hat, herangezogen.

Ziel hierbei ist es, keine ganzheitliche und umfassende Programmtheorie zu rekonstruieren, sondern lediglich jene Annahmen und ihre Beziehungen punktuell zu explizieren, die in den Interviewsituationen als besonders selbstverständlich dargestellt wurden, wobei das Konzept der „Programmtheorie“ der Theoriebasierten Evaluation als theoretischer Rahmen fungieren soll. Dadurch soll den Programmverantwortlichen ein Ausgangspunkt und/oder eine Reflexionsmöglichkeit offeriert werden, um zukünftige Programme und deren Programmtheorien zu konzipieren, die dann davon ausgehend mittels einer Theoriebasierten Evaluationen überprüft werden können. Wichtig ist hierbei, dass unterschieden wird zwischen der Programmtheorie, die durch die Programmverantwortlichen konzipiert wird und denjenigen Annahmen, die die Akteurinnen und Akteure in der Praxis besitzen und nach denen sie die pädagogische Praxis gestalten.

In Abbildung 12 ist die rekonstruierte Programmtheorie abgebildet. Die jeweiligen dargestellten Annahmen beziehen sich auf die in Kapitel 5.4 dargestellten Ergebnisse der qualitativen Interviews. Diese sollen aufgrund dessen im Folgenden nicht detailliert wiedergegeben werden, sondern es wird lediglich die allgemeine Struktur der rekonstruierten Programmtheorie erläutert: Auf oberster Ebene befinden sich die Kontextbedingungen, in denen ENB situiert ist. Einerseits werden konkrete Annahmen hinsichtlich der schulischen Bedingungen sowie des Elternhauses gemacht – insbesondere in Bezug auf vorherrschende oder dadurch bedingte Mängel. All diese Annahmen werden herangezogen, um die Intervention durch das Programm bzw. die Programmziele an sich zu rechtfertigen. Dazu werden einerseits Annahmen zu dem Programm und seinen Zielen gemacht („Programm“) sowie den konkreten Rahmenbedingungen, die das Programm schafft, um seine Ziele zu verwirklichen. Daraus ergibt sich auch ein konkreter Personalbedarf, welches die Programmziele unter den gegebenen Rahmenbedingungen umsetzt, indem es die konkreten Programmangebote konzipiert und realisiert. Dieses Angebot beeinflusst wiederum die Schule sowie die partizipierenden Schülerinnen und Schüler.

Die hier dargestellte Programmtheorie wurde bereits im Rahmen der kommunikativen Validierung den Programmbeteiligten vorgestellt. Aufgrund mangelnder Zeitressourcen und inhaltlichen Priorisierung konnte die rekonstruierte Programmtheorie allerdings nicht wie geplant gemeinsam mit den Programmbeteiligten ausführlich interpretiert und diskutiert werden.

Abbildung 12 Rekonstruierte Programmtheorie ausgehend von den qualitativen Interviews



7. Diskussion

7.1 Summative Evaluation

Im Folgenden sollen die im Laufe dieser Arbeit gewonnenen Erkenntnisse vor dem Hintergrund der Zieldimensionen der summativen Evaluation strukturiert dargestellt werden. Die summative Evaluation bezieht sich dabei zum größten Teil auf die quantitativen Längsschnittdaten, die in Kapitel 4 vorgestellt wurden. Des Weiteren wird sich punktuell auf die qualitativen Daten, die vorliegenden Daten über das Programm, die rekonstruierte Programmtheorie sowie die Hospitationen (siehe Kapitel 3) bezogen, um summative Werturteile hinsichtlich der Angebotsqualität zu treffen. In diesem Zuge wird zunächst betrachtet, ob sich Übereinstimmungen oder Widersprüche zwischen der rekonstruierten Programmtheorie und der Wirkungslogik oder den formulierten Qualitätsindikatoren seitens ENB auffinden lassen, die sich als summatives Bewertungskriterium heranziehen lassen. Allerdings muss bei der Interpretation der hier dargestellten Ergebnisse immer berücksichtigt werden, dass es sich hierbei lediglich um die Interpretation der in den qualitativen Interviews getroffenen Aussagen handelt, wobei anzunehmen ist, dass bestimmte Inhalte aufgrund verschiedener Faktoren (Zeit, Interviewsituation, fehlende Nähe zum Interviewpartner etc.) nicht erwähnt wurden bzw. werden konnten. Anschließend sollen die Fragestellungen, die in Kapitel 3.3 unter die summative Evaluation subsumiert wurden, versucht werden zu beantworten. Abschließend sollen noch weitere Ergebnisse der quantitativen und qualitativen Analyse herangezogen werden, um ergänzend zu der Programmtheorie und den Fragestellungen summative Werturteile in Bezug auf den Evaluationsgegenstand (ENB) zu treffen.

Programmtheorie:

Der Vergleich der Programmtheorie mit der Wirkungslogik lässt sich nur schwer treffen, da in der rekonstruierten Programmtheorie die Wirkbeziehungen lediglich zwischen den Oberkategorien (z. B. Angebot beeinflusst Schülerinnen und Schüler) beschrieben wurden und keine konkreten Kausalbeziehungen auf manifester Variablenebene. Allerdings lassen sich dennoch konkrete Aspekte vergleichen: In der Wirkungslogik wird unter der Spalte „Leistungen“ die Angebotsbeschaffenheit beschrieben, welche ebenfalls in der Programmtheorie rekonstruiert wurde. Das Fachpersonal nannte insbesondere die sozio-emotionale und sprachliche Förderung als Programmziel, abstrahierte allerdings größtenteils von der kreativen und motorischen Förderung. Gleichzeitig wurde eine starke Tendenz zur Einzelförderung oder der Förderung mit reduzierter Gruppengröße geäußert, die ebenfalls in der Wirkungslogik abgebildet ist. Übereinstimmungen zwischen den für die Zielgruppe der Schulkinder angestrebten Wirkungen und der Programmtheorie lassen sich für die Förderung der Sprach- und Lesekompetenz sowie den sozialen Fähigkeiten aufzeigen. Die restlichen aufgeführten Wirkungen wurden von dem Fachpersonal in den Interviewsituationen nicht mehrheitlich angeführt. Dies lässt allerdings nicht den Schluss zu, dass diese Zieldimensionen nicht von dem Fachpersonal angestrebt werden, lässt aber zumindest eine leichte Priorisierung erkennen. Darüber hinaus lässt sich in diesem Kontext noch erwähnen, dass einheitlich die Zieldimension der Beziehung angesprochen wurde, die sich nicht konkret in der Wirkungslogik finden lässt – lediglich in Bezug auf eine „beständige Bezugsperson“. Da die positive und langfristige Beziehung als ein großer Einflussfaktor genannt wurde, liegt es nahe, dies zukünftig in die Wirkungslogik zu integrieren. Des Weiteren lässt sich beobachten, dass das Fachpersonal keine Aussagen dahingehend äußerte, welche die Eltern als Zielgruppe von ENB deklarieren. Bezüglich der Lehrkraft wurden ebenfalls bestimmte Aussagen

vereinzelt getroffen – insbesondere ein erhöhter Kooperationswunsch sowie die Entlastung der Lehrkraft. Allerdings wurde vom Fachpersonal nicht (mehrheitlich) expliziert, inwiefern diese Entlastung oder Kooperation erfolgen soll und welche Wirkungen daraus zu erwarten sind, weshalb dies nicht in die rekonstruierte Programmtheorie mitaufgenommen wurde.

Darüber hinaus lassen sich die von ENB angegebenen Qualitätsindikatoren zum Vergleich heranziehen. Der Indikator der „Kontinuität“ korrespondiert, insbesondere im Kontext der Aussagen zu den langfristigen Beziehungen. Der Indikator der „Wirkungsorientierung“ basiert primär auf den Aussagen zu den quantitativen Diagnostikinstrumenten. Für die anderen Qualitätsindikatoren fällt es schwer, konkrete Aussagen heranzuziehen, wobei diese bis zu einem bestimmten Umfang in den Aussagen des Fachpersonals erkannt werden können. Dies soll allerdings nicht weiter spezifiziert werden.

Fragestellungen der summativen Evaluation:

Die Verbesserungen des Leseverständnisses, welche in **Fragestellung 2.1** untersucht wurden, lassen sich anhand der Daten verschiedenfach beobachten. Einerseits lassen sich deskriptiv kontinuierliche Zuwächse der Gesamtmittelwerte des Leseverständnisses aufzeigen – welcher sich ebenfalls in den Standardabweichungen bemerkbar macht, insofern als über die Schuljahre ein immer geringer werdender Anteil der Standardabweichung in den unterdurchschnittlichen Bereich fällt – andererseits unterscheidet sich inferenzstatistisch lediglich der vierte Messzeitpunkt von allen anderen. Gleichzeitig lässt sich in der vierten Klasse beobachten, dass sich die Standardabweichung nun nicht mehr mit dem unterdurchschnittlichen Bereich überschneidet, sondern mit dem überdurchschnittlichen Bereich. Dies kann als positives Indiz hinsichtlich der Fragestellung gedeutet werden. Zieht man darüber hinaus die Effektstärken heran, die sich aus den Unterschieden zwischen den verschiedenen Messzeitpunkten ergeben, so lässt sich zwischen der ersten und der zweiten Klassenstufe ein Cohens d von ca. 0,5, zwischen der zweiten und dritten Klassenstufe von ca. 0,17 und zwischen der dritten und vierten Klassenstufe von 1,2 beobachten und somit eine mittlere Effektgröße von 0,62 zwischen den Messzeitpunkten über die Jahre. In Bezug auf die Fragestellung lässt sich somit sagen, dass Verbesserungen beobachtbar sind, die – bis auf die zweite auf die dritte Klassenstufe – in großen Effektstärken münden. Ob diese Veränderungen auf ENB zurückzuführen sind, kann allerdings nicht abschließend beantwortet werden, da aufgrund des Evaluationsdesigns dieser Arbeit keine Kausalaussagen getroffen werden können.

Darüber hinaus wurde die **Fragestellung 2.2** anhand der SELLMO-Testergebnisse ausgewertet. Hierbei wurden keine inferenzstatistisch signifikanten Unterschiede beobachtet, allerdings weisen die deskriptiven Veränderungen der einzelnen motivationalen Orientierungen jeweils in eine positiv zu interpretierende Richtung hinsichtlich der Veränderungen der Mittelwerte – wobei die Reduktion der Ausprägung der Annäherungs-Leistungzielorientierung auf Grundlage der zugrundeliegenden Theorie nicht sinnvoll interpretiert werden kann. Betrachtet man darüber hinaus die Veränderungen der Standardabweichungen lässt sich bei der Lernzielorientierung eine Verringerung der Heterogenität feststellen, wodurch sich einerseits ein kleinerer Bereich der Standardabweichung im unterdurchschnittlichen Bereich befindet, aber andererseits auch die obere Grenze der Standardabweichung in der vierten Klasse einen kleineren Wert aufweist als in der dritten Klasse. Diese Verringerung der oberen Grenze lässt sich ebenfalls bei der Vermeidungs-Leistungzielorientierung sowie bei der Arbeitsvermeidungsorientierung

erkennen, wobei die unteren Grenzen relativ gleich bleiben. Bei der Annäherungs-Leistungszielorientierung vergrößert sich die Heterogenität in der vierten Klassenstufe im Vergleich zur dritten Klassenstufe. Wie dies im Hintergrund der nachzuweisenden Wirkung von ENB zu interpretieren ist, lässt sich hier nicht abschließend klären. Positiv scheint die unterdurchschnittliche Ausprägung der Vermeidungs-Leistungszielorientierung der Kinder zu sein, worauf das Fachpersonal im Rahmen der kommunikativen Validierung äußerte, dass es ebenfalls einen Anstieg der Fragefrequenz in der vierten Klassenstufe bei den Kindern beobachtete, in anderen Worten: Die Kinder fragten in der vierten Klasse viel häufiger nach und zeigten somit eigene Wissens- oder Kompetenzlücken.

Weitere summative Werturteile:

Im Folgenden werden insbesondere die Aussagen der Programmbeteiligten herangezogen, um summative Werturteile zu approximieren. Allerdings sollte hier berücksichtigt werden, dass hierbei lediglich deren Aussagen und nicht die tatsächlich beobachtete Praxis zur Bewertung herangezogen werden. Gleichzeitig werden nur jene Aspekte dargestellt, die ein positives Werturteil hinsichtlich ENB zulassen, da identifizierte Defizite oder Verbesserungsmöglichkeiten in Kapitel 7.4 dargestellt werden; allerdings wird das positive Werturteil partiell differenziert und kritisch hinterfragt.

Zunächst lässt sich anführen, dass sich die Interviewten sehr zufrieden mit dem Programm äußerten und darüber hinaus einen wertschätzenden Umgang zwischen den Programmbeteiligten beschreiben. Gleichzeitig wird das Programm einstimmig als wirksam und sinnvoll wahrgenommen. Diese positiv geäußerte Einstellung des Fachpersonals gegenüber dem Programm lässt sich positiv für das Programm bewerten. Darüber hinaus sollen noch weitere Punkte angeführt werden, die sich insbesondere auf Erkenntnisse aus StEG beziehen:

Passung mit Bedürfnissen und Interessen der Schülerinnen und Schüler:

StEG beschreibt, dass die positive Angebotsqualitätsbewertung zu besseren Leistungstestergebnissen führt, wenn eine Passung zwischen dem Angebot sowie den Bedürfnissen und Interessen der Schülerinnen und Schüler vorliegt (Jugendinstitut, 2016). Anhand der Aussagen des Fachpersonals lässt sich eine starke Orientierung der methodisch-didaktischen Gestaltung der Unterrichtseinheiten an den Schülerbedürfnissen und -interessen beobachten, wonach dies als positiv bewertet werden kann. Insbesondere der starke Fokus des Programms auf den sozio-emotionalen Bereich gilt nach StEG als ein Faktor, dass die Angebotsqualität durch die Schülerinnen und Schüler als positiv bewertet wird (Jugendinstitut, 2016).

Freiwilligkeit:

Zieht man die Erkenntnisse aus StEG an anderer Stelle heran, so lässt sich die fehlende Freiwilligkeit hinsichtlich der Angebotsteilnahme seitens der Schülerinnen und Schüler kritisch beleuchten (Jugendinstitut, 2016): Wie in Kapitel 2.4 beschrieben, geht lediglich die freiwillige Teilnahme mit statistisch signifikanten Kompetenzzuwächsen einher. Allerdings bemerken die Autorinnen und Autoren von StEG darüber hinaus an, dass bei Kindern mit Migrationshintergrund bei kontinuierlicher Nutzung der Angebote – unabhängig der Freiwilligkeit – Zuwächse der sozialen Kompetenzen zu beobachten sind (Jugendinstitut, 2016). Vor diesem Hintergrund und der hohen Anzahl der teilnehmenden Kinder mit Migrationshintergrund lässt sich für ENB schlussfolgern, dass das Moment der Freiwilligkeit

eventuell jene Kinder mit Migrationshintergrund, die das Angebot nicht genutzt hätten, somit nicht gefördert hätte und deshalb kann dies zumindest partiell im Hinblick auf die Zielgruppe als positiv gewertet werden.

Systematizität und Kooperation mit Wissenschaft:

StEG beschreibt darüber hinaus, dass außerunterrichtliche Angebote systematisch geplant und wissenschaftlich begleitet werden sollen, um die Qualität der Angebote zu gewährleisten (Jugendinstitut, 2019). Ausgehend von Kapitel 3, lässt sich das Programm von ENB als systematisch beschreiben, insofern als es Programmziele, eine Wirkungslogik und (systematisch) aufeinander abgestimmte Förderangebote umfasst. Gleichzeitig lässt sich auch ein systematischer und langfristiger Austausch und Begleitprozess seitens der Wissenschaft feststellen. Neben den bisherigen summativen und formativen Evaluationen, die von wissenschaftlichen Akteurinnen und Akteuren durchgeführt wurden, wurde auch die Programmkonzeption wissenschaftlich begleitet. Nach Aussage der Programmleitung wurden die Erkenntnisse der Evaluationen insbesondere zur Reflexion über das Programm herangezogen und punktuell in konkrete Veränderungen überführt. Somit lässt sich konkludieren, dass sowohl die Programmkonzeption als auch die Programmdurchführung systematisch erfolgte und wissenschaftlich begleitet wurde und das Programm dementsprechend die Anforderungen, die StEG formuliert, approximiert.

Kooperation mit Lehrkräften:

StEG beschreibt des Weiteren die Wichtigkeit der (multiprofessionellen) Kooperation zwischen dem Fachpersonal und den schulischen Akteurinnen und Akteuren, um individuelle Förderung erfolgreich umsetzen zu können (Jugendinstitut, 2019). Mit der gegebenen Datenlage erscheint es schwierig, hier zu einem Urteil zu kommen. Andererseits konkretisiert StEG auch nicht, welche Kooperationsprozesse in welcher Ausprägung und Intensität gegeben sein müssen, um einen Einfluss auf die Qualität zu haben. Da das Fachpersonal den Kooperationsprozess mit der Lehrkraft und den weiteren schulischen Akteurinnen und Akteuren noch ausbaufähig sieht und an manchen Stellen insbesondere Informationen über die Sichtweise der Lehrkraft auf bestimmte Schülerinnen und Schüler gewünscht wird, kann abschließend geurteilt werden, dass zwar Kooperationsprozesse vorhanden sind, diese aber noch ausbaufähig sind.

Räume und Zeit:

Zur Kooperation gehört darüber hinaus das Schaffen von räumlichen Strukturen und zeitlichen Ressourcen, in denen die Angebote sowie die Kooperationsprozesse situiert werden (Jugendinstitut, 2019). Das Fachpersonal beschreibt die Vergabe der Räume und Zeiten im Rahmen von Aushandlungsprozessen, die es mit den schulischen Akteurinnen und Akteuren realisiert, wobei hierbei teilweise Schwierigkeiten beschrieben werden und das Fachpersonal die Wahrnehmung äußert, dass die schulischen Ziele häufiger über den eigenen Programmzielen gestellt werden. Zum Teil wurde ebenfalls vom Fachpersonal ausgesagt, dass die zur Verfügung gestellten Räume wenig Raum für Gestaltung lassen und das eigene Angebot den Raum nicht so miteinbeziehen kann, wie vom Fachpersonal erwünscht. Auf der anderen Seite wurden keine existenziellen Mängel bekundet, wie fehlende Räume oder Zeiten, sodass das Programm wie geplant durchgeführt werden konnte. In Anbetracht dessen lässt sich schlussfolgern, dass die zur Programmdurchführung notwendigen räumlichen und

zeitlichen Ressourcen vorhanden sind, hierbei allerdings noch Verbesserungspotenzial besteht.

Verzahnung der ganztägigen Angebote:

Nach Aussagen der Kultusministerkonferenz (KMK, 2023) in Anlehnung an die Erkenntnisse von StEG wird insbesondere das Potenzial der Verzahnung der ganztägigen Angebote sowie die multiprofessionelle Kooperation größtenteils nicht realisiert. Betrachtet man ausgehend von diesem Anspruch das Programmangebot und die Ausrichtung auf das schulische Angebot, so lässt sich zunächst konzeptionell urteilen, dass sich dies als inhärenter Programmbestandteil auffassen lässt: Neben der Unterstützung der Lehrkraft bei der Aufgabenbetreuung der Kinder zielen die Programmziele zum Teil darauf ab, die Kinder in den schulischen Fach- und Grundkompetenzen wie Lesen, Sprechen, Interesse an schulischen Inhalten aber auch Motivation zu fördern. Kritisch lässt sich allerdings betrachten, inwiefern die Programmangebote mit den schulischen Verantwortlichen kontinuierlich abgestimmt und aufeinander angepasst wurden – eventuell auch, inwiefern sich die schulischen Ganztagsangebote an die Programmangebote aufgrund von Interaktions- und Kommunikationsprozessen zwischen den Beteiligten angepasst haben. Da die Kommunikation mit den schulischen Beteiligten nach Aussage des Fachpersonals nicht institutionalisiert wurde und dahingehend keine Aussagen gemacht wurden, lassen sich diese kontinuierlichen, gegenseitigen Ausrichtungsprozesse zumindest nicht als gegeben annehmen.

Qualitätssicherungs- und -entwicklungsprozesse:

Nach KMK (2023) ist ein Qualitätsmerkmal das Sichern der Qualität der Ganztagsschulangebote und das Überprüfen ihrer Wirkung. Auf Grundlage der Aussagen des Fachpersonals lässt sich konkludieren, dass Qualitätssicherungsprozesse auf Programmebene kaum realisiert werden konnten, da einerseits eine zugehörige Stelle seitens GmS abgebaut wurden und andererseits das Fachpersonal sich einheitlich als ressourcentechnisch und kapazitiv nicht ausgestattet beschreibt, um Qualitätssicherungsprozesse realisieren zu können. Hierbei sollte künftig zwischen Angebots- und Programmqualitätssicherungsprozessen differenziert werden, um zu einem abschließenden Urteil zu kommen, was nicht in dieser Arbeit erfolgte. Aufgrund der großen Freiheiten lässt sich allerdings konzeptionell die Möglichkeit der Angebotsqualitätssicherung und -entwicklung skizzieren, die allerdings gänzlich in die Hände der jeweiligen Fachkraft gelegt wird.

Programmstruktur:

Anhand der Aussagen des Fachpersonals lässt sich deduzieren, dass das Programm das Schaffen von motivierenden, inklusiven und bedürfnisgerechten Lern- und Freizeitgelegenheiten innerhalb der außerschulischen Angebote sowie die Kooperation zwischen den außerschulischen und schulischen Akteursgruppen ermöglicht. Hierbei lässt sich aber noch ein großer Ausbaubedarf feststellen.

7.2 Formative Evaluation

Im Folgenden sollen die im Rahmen der formativen Evaluation formulierten Fragestellungen kurz und bündig beantwortet werden. Für detailliertere Informationen wird auf Kapitel 5 verwiesen.

1.1 Welche Verläufe des Leseverständnisses lassen sich über die vier Jahre beobachten?

In den Leistungsverläufen des Leseverständnisses in Kapitel 4.5.3 lassen sich deskriptiv unterschiedliche Verläufe beobachten. Am prägnantesten fällt auf, dass beinahe alle Kinder einen Anstieg der T-Werte von der dritten auf die vierte Klassenstufe aufzeigen. Darüber hinaus lassen sich teilweise Zickzackverläufe beobachten, bei denen die Schülerinnen und Schüler zunächst einen Zuwachs und anschließend einen Abstieg aufweisen; insbesondere von der zweiten auf die dritte Klassenstufe. Dies könnte eventuell darauf zurückzuführen sein, dass es von der zweiten auf die dritte Klasse einen Wechsel der Klassenlehrkraft gab.

1.2 Wie gestaltet das Fachpersonal die Unterrichtsstunden und was ist ihm dabei wichtig?

Das Fachpersonal äußerte einheitlich, dass die Planung einer Unterrichtsstunde häufig zunächst ausgehend des zu verwendenden Materials geschieht, woran sich die methodisch-diaktische Integration des Materials in das Angebot anschließt. Als besonders wichtig wurde die Passung des Materials mit den Interessen und Bedürfnissen der Kinder beschrieben, die insbesondere aufgrund der positiven und langfristigen Beziehung mit den Kindern besser diagnostizierbar sind. Für die Programmverantwortlichen lässt sich hierbei ableiten, inwiefern das Fachpersonal bei der Materialwahl stärker unterstützt werden könnte.

1.3 Welches Verständnis hat das Fachpersonal von den Programmzielen?

Diese Fragestellung korrespondiert stark mit der rekonstruierten Programmtheorie, weshalb auf Kapitel 6 verwiesen wird.

1.4 Wie nimmt das Fachpersonal die Kooperation mit und die Integration in der Schule wahr?

Das Fachpersonal betonte stark einen gemeinsamen Lernprozess, den es mit den schulischen Beteiligten durchlaufen hatte. Die konkreten Kooperationsprozesse finden lediglich im informellen Rahmen statt. Gleichzeitig wird die eigene Stellung in der Schule als „Randgeschäft“ wahrgenommen und fehlende Kommunikations- und Austauschprozesse beschrieben.

1.5 Welche Einstellung hat das Fachpersonal gegenüber dem Programm?

Wie bereits erwähnt ist die Einstellung des Fachpersonals gegenüber dem Programm einstimmig positiv.

1.6 Welches Verständnis hat das Fachpersonal in Bezug auf seine individuelle Rolle im Programm?

Abhängig des jeweiligen biographischen und professionellen Hintergründen sowie der konkreten Vorgaben seitens ENB äußerte das Fachpersonal teils unterschiedliche Rollenverständnisse. Allerdings wurde einhellig bekundet, dass der Prozess der Rollendefinition über die Förderzeit zu Beginn als schwieriger beschrieben wurde; insbesondere in dem Umgang mit den gewährten Freiheiten.

1.7 Wie trifft das Fachpersonal diagnostische Urteile und Wirksamkeitsurteile?

In den CIT-Interviews ließen sich verschiedenartige Strategien und Präferenzen erkennen, wie das Fachpersonal diagnostische und Wirksamkeitsurteile fällt. Größtenteils wurden allerdings konkrete Interaktionssituationen sowie die Beziehung zu dem Kind als primäre Quellen diagnostischer Informationen beschrieben, die darüber hinaus Aussagen über die Wirksamkeit der eigenen Intervention zuließen.

1.8 Wie lässt sich das Programm evidenzbasierter gestalten?

Hinsichtlich der Integration der Ergebnisse der Diagnostiktests äußerte sich das Fachpersonal ambivalent, da diese teilweise als konkurrierend zu ihren eigenen alltäglichen diagnostischen Urteilen und Einstellungen aufgefasst wurde. Darüber hinaus erwähnt das Fachpersonal, dass es sich ressourcentechnisch nicht ausgestattet sieht, konkrete evidenzbasierte Qualitätssicherungs- und -entwicklungsprozesse zu realisieren. Inwiefern das Fachpersonal selbst evidenzbasierte Entscheidungen fällt, wurde nicht gefragt.

1.9 Welche Einstellung hat das Fachpersonal gegenüber quantitativen diagnostischen Instrumenten und deren Integration in die pädagogische Praxis?

Hierzu äußerte sich das Fachpersonal ebenfalls ambivalent. Einerseits wurde die Validität des eigenen diagnostischen Urteils angeführt – insbesondere aufgrund des „direkten“ Kontaktes mit den Kindern – andererseits jedoch wurde den quantitativen Diagnostikinstrumenten eine mögliche Funktion als Korrektiv des eigenen diagnostischen Urteils zugesprochen. Die Integration in die pädagogische Praxis wird speziell aufgrund des Zielkonflikts der Erhebungssituation geprägt, da abgewägt wird, ob das Kind den Test bearbeitet oder in derselben Zeit die Förderung durch das Fachpersonal erhält.

1.10 Welche Strategien wendet das Fachpersonal im pädagogischen Umgang mit quantitativen diagnostischen Instrumenten an?

Es ließ sich insbesondere beobachten, dass einerseits die Ergebnisse der Diagnostiktests im Vergleich mit dem eigenen diagnostischen Urteil bewertet wurden und andererseits dass die Auseinandersetzung zu Interpretations- und Reflexionsprozessen über die eigene Angebotsgestaltung sowie die Programmziele führte. Gleichzeitig wurde von einem Teil des Fachpersonals ein großer Zweifel an der Validität und Sinnhaftigkeit der Diagnostiktests geäußert. Gleichzeitig hängen Messbarkeitsüberlegungen der im CIT-Interview vorkommenden latenten Konstrukten stark von den individuellen Erfahrungen des Fachpersonals ab.

7.3 Limitationen der Arbeit

Im Folgenden sollen die Limitationen der Arbeit dargestellt werden. Diese beziehen sich einerseits auf die quantitativen und andererseits auf die qualitativen Analysen.

Quantitative Analyse:

Zunächst lässt sich anmerken, dass die Auswahl der Datenbasis, die für die quantitative Analyse dieser Arbeit herangezogen wurde, vornehmlich nach Verfügbarkeitskriterien geschah. So lagen darüber hinaus noch Daten aus einem Beobachtungsbogen der Jahrgangsstufen 1-3 vor (nicht jedoch für die vierte Jahrgangsstufe), welcher die sozio-emotionalen Kompetenzen der Kinder mittels Fremdbeobachtung seitens der Programmleitung einschätzte sowie ein Elternfragebogen, welcher allerdings nicht erhoben wurde, der insbesondere sozio-demographische Informationen zu den einzelnen Kindern für die Analyse zugänglich gemacht hätte.

Des Weiteren bestand eine große Abhängigkeit der Erhebungssituationen der quantitativen Diagnostiktests und deren Auswertung von der Person der Programmleitung, da diese einerseits den Ort und Zeitpunkt der Testung sowie den tatsächlichen Testablauf kontrollieren konnte. Hierdurch werden – zumindest unbewusste – Einflussmöglichkeiten auf die Testergebnisse ermöglicht und die Validität der Testergebnisse in das Vertrauen der Programmleitung übergeben, die dadurch in einen Rollenkonflikt gerät, da die Testergebnisse die Güte der eigenen Arbeit einschätzen sollen.

Darüber hinaus lässt sich limitierend anmerken, dass die vorliegenden Stichproben vergleichsweise klein sind und es keine Kontrollgruppe gibt, an denen die Ergebnisse relativiert werden können. Gleichzeitig lässt sich die Vergleichbarkeit der Normierungsstichproben kritisieren, einerseits aufgrund der möglichen Auswirkungen der Corona-Pandemie als Periodeneffekt sowie der große Anteil an Kindern mit Migrationshintergrund in der Klasse. Somit stellt sich die Frage, inwiefern die Aussagen über einen „Normalbereich“ und die zu erwartenden Leistungszuwächse, die sich durch die Normierungsstichproben ergeben, auf diese Stichprobe übertragbar sind.

Bezüglich der Datenanalyse bestehen die Limitationen darin, dass einerseits die ELFE-Testergebnisse lediglich mit den Gesamtwerten berechnet und nicht die einzelnen Untertests analysiert wurden, die eventuell andere Aussagen zugelassen hätten. Gleichzeitig lassen sich noch weiterführende Analysen skizzieren, die ebenfalls berechnet hätten können,

um die vorhandene Datenbasis optimal auszunutzen, wie z. B. Interaktionen zwischen dem Leseverständnis, der Motivation und den Noten.

Qualitative Analyse:

Im Kontext der qualitativen Analyse lassen sich ebenfalls einige Limitationen anführen. Einerseits stellt die zur Verfügung stehende Zeit für die Interviews eine große Limitation dar, da hierdurch bei der Leitfadenkonstruktion nur wenige Inhalte in der gewünschten Tiefe integriert werden konnten. Insbesondere die CIT-Interviews hätten von einer längeren Interviewdauer profitieren können. Darüber hinaus führte die Entscheidung der Anonymisierung der erhobenen Daten dazu, dass keine personenbezogenen Aussagen getroffen werden konnten, was den Detaillierungsgrad der quantitativen Analyse reduzierte und mögliche, bereichs- und personenbezogene Erkenntnisse exkludierte. Letztlich lässt sich noch anmerken, dass die relative Unbekanntheit des Autors dieser Studie seitens des Fachpersonals sowie der Abhängigkeit der Aussagen von der Programmbewertung dazu führen hätte können, dass bestimmte angefragte Informationen nicht oder modifiziert geäußert wurden.

7.4 Empfehlungen

Im Folgenden sollen nun die Ergebnisse aus der summativen und formativen Evaluation herangezogen werden, um für die Programmverantwortlichen konkrete Empfehlungen auszusprechen, die auf zukünftige Programme übertragen werden könnten.

Qualitätssicherung und –entwicklung:

Hinsichtlich der Qualitätssicherungs- und -entwicklungsprozesse erscheint es empfehlenswert, eine eigene Stelle zur Qualitätssicherung und –entwicklung zu schaffen und dies strukturell im Programm zu verankern. Es scheint nicht realistisch, dies als weiteren Aufgabenbereich der Programmleitung aufzuerlegen, da nach Einschätzung des Autors dieser Studie die Programmleitung einerseits stark in der Angebotsdurchführung involviert ist und andererseits das Hauptaugenmerk auf dem Austausch mit den schulischen Akteurinnen und Akteuren und dem Fachpersonal liegen sollte sowie des Managements der Programm- und Angebotsabläufe; insbesondere die Sicherstellung, dass das Fachpersonal die geeigneten Rahmenbedingungen hat, um ihr Angebot konzipieren und realisieren zu können. Hierbei ließe sich auch vorstellen, bestimmtes Erfahrungswissen bezüglich der Kursadministration oder anderen angebotsbezogenen Erfahrungen im Rahmen eines Wissensmanagements zu dokumentieren, um die Einarbeitung neuen Personals bei Personalwechseln zu fördern.

Evidenzbasierte pädagogische Entscheidungen:

Aufgrund der Verwendung der quantitativen Diagnostikinstrumente zu Evaluationszwecken ist es vorstellbar, die Testergebnisse ebenfalls für pädagogische Entscheidungen heranzuziehen.

Umgang mit Programmtheorie:

Mittels der Rekonstruktion der Programmtheorie durch diese Arbeit wurden teilweise implizite Annahmen des Programms aufgedeckt und somit dem Diskurs zugänglich gemacht. Als weiteren, nächsten Schritt gilt es zu eruieren, ob die hierbei rekonstruierten Inhalte den

Erwartungen der Programmverantwortlichen entsprechen und inwiefern hierfür auf empirische Evidenzen aus der sozialwissenschaftlichen Forschung zurückgegriffen werden kann. Hierbei erscheint es sinnvoll, die Programmtheorie seitens der Programmverantwortlichen weiter zu detaillieren und gezielt mit sozialwissenschaftlichen Erkenntnissen zu fundieren. Allerdings ist hierbei zentral, welche empirischen Erkenntnisse wie herangezogen werden: Es muss immer berücksichtigt werden, dass das Angebot des Programms vornehmlich aus "sozialen Interaktionen" besteht, die hochgradig von den jeweiligen Akteurinnen und Akteuren und den Kontextfaktoren geprägt sind. Dies sollte bei der Interpretation der empirischen Evidenzen und der Übertragung auf die eigene Programmtheorie berücksichtigt werden. Ziel der Festlegung der Programmtheorie ist es, den Weg für zukünftige Evaluationen zu bereiten, die dadurch die Evaluation des Programms theoriebasiert als Überprüfung der Programmtheorie durchführen können. Gleichzeitig sollte die Programmtheorie immer an den Kontext angepasst werden, somit ständig überarbeitet werden.

Institutionalisierung an Schule:

Die Erfahrungen dieser Phase sollten zu konkreten Überlegungen führen, welche Strukturen an der Schule etabliert werden sollten, um die Angebotsqualität zu steigern bzw. das Programmangebot so umsetzen zu können, wie vorgesehen. Die daraus gewonnenen Inhalte – insbesondere bezüglich der Kooperation mit den schulischen Beteiligten – sollten dokumentiert werden. Für zukünftige Programme lässt sich vorstellen, dass bestimmte Kommunikations- und Austauschprozesse abgeklärt und als „Regelwerk“ schriftlich festgehalten werden sollte, worauf sich bei Bedarf bezogen werden kann. Das Augenmerk sollte besonders auf der Einbettung des Programms in die schulischen Kommunikationsabläufe sowie die Integration des Fachpersonals in das Lehrerkollegium liegen; mit der Zielvorstellung, dass das Fachpersonal als gleichwertige Mitglieder des Teams angesehen werden. Ebenfalls sollten seitens der Schule und des Programms Zeitfenster eingeplant werden, die den Austausch zwischen schulischen Akteuren und Programmteilnehmenden ermöglichen, wobei darauf geachtet werden sollte, dass hierdurch keine zusätzlichen Belastungen entstehen.

Personalentwicklung:

Es empfiehlt sich, gezielt Fortbildungsmaßnahmen für das Fachpersonal mit in die Budgetierung und Planung zu integrieren. Dies könnte u. a. von der Stelle für Qualitätssicherung und -entwicklung übernommen werden, die qualitätsbezogene Fort- und Weiterbildungen plant und koordiniert. Gleichzeitig sollte darauf geachtet werden, welche konkreten Kompetenzen das Fachpersonal benötigt, um die Angebote qualitativ zu gestalten und die Programmziele zu erfüllen.

Qualitätsstandards:

Darüber hinaus empfiehlt es sich, für die einzelnen Förderbereiche gezielt Qualitätsstandards festzulegen und Qualitätsentwicklungsmaßnahmen anzuwenden. Dies sollte einerseits unter der Wahrung der professionellen Freiheit des Fachpersonals und andererseits unter Einbezug des Fachpersonals bei der Festlegung der Qualitätsstandards geschehen.

Unterrichtshospitationen:

Eventuell bietet es sich an, kollegiales Feedback in Form von Unterrichtshospitationen und anschließenden Feedbackgesprächen strukturell zu verankern. Da die Fachkräfte große

Freiheiten in ihren Rollen und der Angebotsgestaltung besitzen, werden dadurch Feedbackprozesse geschaffen und das Fachpersonal kann von den multiprofessionellen Hintergründen der anderen profitieren.

Einheitlicher Plan für Umgang mit quantitativen Diagnostikdaten:

Die quantitativen Diagnostiktests eignen sich insbesondere, um auffällige Eigenschaftsausprägungen bei den Kindern zu identifizieren und mögliche Ursachen heranzuziehen. Darauf aufbauend lassen sich gezielt Fördermaßnahmen konzipieren, deren Wirksamkeit anhand des wiederholten Einsatzes der Diagnostikinstrumente überprüft werden kann. Darüber hinaus bietet sich ein konzertiertes Vorgehen über alle Förderangebote hinweg an, insofern als mit den Diagnostiktests oder anderweitig identifizierten Förderbedarfen das jeweilige Kind in den verschiedenen Unterrichtsangeboten seitens des Programms bedarfsspezifisch gefördert wird. Bspw. bei diagnostizierten Defiziten im Wortverständnis bei einem Kind bietet es sich an, dass jede Fachkraft in ihrem speziellen Unterrichtsangebot dem Kind verschiedene Zugänge anbietet, welche das Wortverständnis steigern sollen. Hierfür müsste abgeklärt werden, welche Ressourcen zur Verfügung stünden und wie viele Kinder zu einem Zeitpunkt gefördert werden könnten. Gleichzeitig muss ein Zeitrahmen festgelegt werden, der optimalerweise mit einem Diagnostiktest zur summativen Evaluation endet. Dies beinhaltet gleichfalls spezifische Kompetenzen und Ressourcen, die Testergebnisse aufzubereiten und damit umgehen zu können. Allerdings kommt es durch die Auseinandersetzung mit den Diagnostiktests auch zu Reflexionsprozessen und zur Argumentation mit eigenen subjektiven Theorien. Hierdurch ließe sich dies als geeigneter Anlass konzipieren, um auch subjektive Theorien zwischen dem Fachpersonal auszutauschen und in einen (professionellen) Diskurs zu kommen.

Wissenschaftliche Begleitung zur Interpretation der Diagnostiktests und Ableitung von Fördermaßnahmen:

Es erscheint ratsam, zur allgemeinen Evaluation des Programms und zur Generierung von Empfehlungen auf entsprechende Experten zurückzugreifen, da insbesondere die Interpretation der Ergebnisse ausreichende psychometrische Kenntnisse voraussetzt. In diesem Zuge erscheint es sinnvoll, einen Katalog an evidenzbasierten Fördermaßnahmen festzulegen, auf den das Fachpersonal bei Vorliegen bestimmter Förderbedarfe zurückgreifen kann. Hierbei sollte ebenfalls längerfristig eine Ansprechperson zur Verfügung gestellt werden, die vertieftes psychometrische und Testkenntnisse besitzt und bei aufkommenden Fragen zur Verfügung steht. Es erscheint darüber hinaus sinnvoll, die durch die unterschiedlichen Tests gewonnenen Informationen aufeinander zu beziehen und daraus individuelle Förderprofile zu erstellen. Diese Idee wurde bereits durch das Fachpersonal geäußert und würde somit teilweise auf Akzeptanz stoßen. Um den diagnostischen Prozess umfassend zu gestalten und mit standardisierten Instrumenten zu unterstützen, empfiehlt es sich, Informationen hinsichtlich folgender kindlicher Eigenschaften zu erheben (vgl. Spinath et al., 2012):

- **Fähigkeiten:** z. B. Intelligenz und (fach-)spezifische Leistungen.
- **Motivation:** z. B. Lern- und Leistungsziele.
- **Emotionen:** z. B. Prüfungsängstlichkeit, Depressionen.
- **Kognitive Überzeugungen und Denkschemata:** Fähigkeitskonzept, Attributionen
- Darüber hinaus sollten **familiäre und soziale Umgebungsbedingungen** exploriert und in den Interpretationsprozess integriert werden

Umsetzung ELFE:

Für den ELFE-Test existiert für die Klassenstufen 1-3 eine Kurzversion, die auf die Aufgaben zum Textverständnis verzichtet, wodurch der zeitliche Aufwand reduziert werden kann. Dies bietet sich dann an, wenn der Textverständnisteil noch zu schwierig für das Kind ist. Es bietet sich ebenfalls an, die Normen für Kinder mit Migrationshintergrund heranzuziehen nach: Lenhard, A. & Lenhard, W. (2023). Diagnostik von Lesestörungen mit ELFE II bei Kindern mit Migrationshintergrund (inklusive adaptierter Normen, 2. aktualisierte Auflage). Psychometrica.

Es könnte auch in Betracht gezogen werden, die Computerform des Tests zu nutzen, da hierdurch Anwendefehler minimiert werden und noch mehr Zeit bei der Erhebung und Aufbereitung eingespart werden kann – insbesondere da GmS über Tablets verfügt, die hierfür verwendet werden können. In der Computerform werden darüber hinaus die meisten Kennwerte berechnet, wodurch sie sich besser für eine differenzierte Einzeldiagnostik eignen – auch für die Diagnostik von Lese-Rechtschreibstörungen. Gleichzeitig werden weniger Anforderungen an das Personal bei der Durchführung des Tests gestellt.

Konsequenzen der Verwendung des SELLMO-Tests:

Die Verwendung des SELLMO-Tests impliziert, dass das Programm das Ziel verfolgt, bei den Schülern eine Lernzielorientierung zu schaffen und die anderen Zielorientierungen möglichst abzubauen. Aufgrund dessen erscheint es empfehlenswert, besonders darauf zu achten, dass das Personal über die notwendigen theoretischen Kenntnisse verfügt, um gezielt die jeweiligen motivationalen Orientierungen zu fördern. Das sollte dabei einerseits aktiv durch das Fachpersonal in die Angebotskonzeption, -durchführung und -reflexion integriert und andererseits struktureller Bestandteil des Programms werden. Darüber hinaus sollte überlegt werden, ob die anvisierte Zielgruppe hinreichend über Sprach- und Lesekenntnisse verfügt, um den SELLMO-Test bearbeiten zu können.

Allgemeine Zielsetzung des Programms:

Um Wirksamkeitsnachweise durch quantitative Tests sinnvoll in das Programm einzubetten, sollten die dahinterstehenden Konstrukte systematisch in die Programmtheorie aufgenommen und das Programmkonzept strukturell demnach gestaltet werden. Andererseits kann nicht gewährleistet werden, dass die getesteten Konstrukte tatsächlich im Unterricht zielgerichtet und bedarfsorientiert vermittelt werden. Beschreibt nicht nur die allgemeine Ausrichtung auf Leseverständnis oder Motivation, sondern auch auf die Unterkonstrukte wie Wort-, Satz- und Textverständnis sowie Lernzielorientierung etc. Gleichzeitig sollte sichergestellt werden, dass das Personal über das notwendige konstruktbezogene, theoretische Wissen und entsprechendes Interventionswissen/-kompetenz verfügt, i. S. v. geeignete Interventionen abhängig der Ausprägung des jeweiligen Konstruktes wählen bzw. konzipieren und durchführen zu können. Es empfiehlt sich hier, mit den Programmbeteiligten in den diskursiven Austausch zu treten und abzuwägen zwischen den gewährten Freiheiten und der Ausrichtung auf konkrete Programmziele. Allerdings sollte hier behutsam und möglichst verständnisvoll umgegangen werden, da hinsichtlich einer allgemeinen Zielausrichtung und der Operationalisierung jener Ziele bei einem kleinen Teil des Fachpersonals aversive Einstellungen vorherrschen. Im Allgemeinen lässt sich dies als „Evidenzbasierung“ beschreiben, insofern als das Fachpersonal zunächst wissenschaftliche Theorien aneignet und diese und damit verbundene wissenschaftliche Interventionen in das Unterrichtsgeschehen integriert.

Umgang mit Freiheit:

Die dem Fachpersonal hinsichtlich der Angebotsgestaltung gegebene Freiheit wird einheitlich als sehr positiv und wertschätzend aufgefasst. Allerdings erwähnte das Fachpersonal auch, dass hierdurch Hürden entstehen, die es gilt, erfolgreich zu überwinden. In diesem Kontext wurde allerdings einheitlich beschrieben, dass diese Hürden zum Zeitpunkt der Interviews überwunden wurden. Neben den durch das Fachpersonal erwähnten Hürden der eigenen Rollendefinition (vermittelnd zwischen eigenen Ansprüchen, denen des Programms, der Schüler und der Schule) sowie die methodisch-didaktische Konzeption des Unterrichts insbesondere vor dem Hintergrund der räumlichen und situativen Rahmenbedingungen, lassen sich noch weitere Hürden skizzieren, die nicht erwähnt wurden, aber dem Autor dieser Studie plausibel erscheinen:

Einerseits lässt sich vorstellen, dass mit ansteigender Anforderungen an das Fachpersonal hinsichtlich der Förderung konkreter latenter Konstrukte (wie Leseverständnis, Lernzielorientierung etc.) die Zeit und den Ressourcenbedarf (Zeit, Lektüre anspruchsvoller Studien etc.) der Vorbereitung des Unterrichts erhöht. In der jetzigen Ausgestaltung scheint es eher so zu sein, dass das Unterrichtsangebot nach individuellen und situativen Maßstäben gestaltet wird, was hochgradig von der individuellen Professionalität und Motivation abhängig ist. Die Unterrichtsinhalte und -methoden werden nach Aussagen des Fachpersonals einerseits durch die Bezugnahme auf wahrgenommene kindliche Interessen und Bedürfnisse ausgewählt und andererseits durch situative Faktoren und individuelle Präferenzen bestimmt, insbesondere das Finden geeigneten Unterrichtsmaterials scheint eher durch situative Faktoren beeinflusst zu werden. Darüber hinaus müsste erörtert werden, inwiefern das Unterrichtsangebot bestimmten Qualitätsanforderungen seitens des Programms und seitens der Wissenschaft gerecht wird. Das Einhalten dieser Qualitätsanforderungen würde darüber hinaus die Unterrichtsvor- und -nachbereitung ebenfalls in die Länge ziehen. Die Programmverantwortlichen sollten sich aufgrund dessen bewusst sein, dass mit großer gewährten Freiheit hinsichtlich der Programmausgestaltung geringere Steuerungsmöglichkeiten einhergehen und somit immer die Möglichkeit besteht, dass die Ziele des Angebots (bzw. die Ziele des Fachpersonals) und die des Programms nicht konvergieren.

Überlegung hinsichtlich der Angebotsgestaltung:

Darüber hinaus scheint es sinnvoll, anhand der Ergebnisse dieser und vorangegangener Evaluationen auf konzeptioneller Ebene zu reflektieren, welches Angebotsformat am geeignetsten ist, um die Programmziele umzusetzen. Wenn das bestehende Angebotsformat beibehalten werden soll, erscheint es ratsam, die Zielsetzung des Programms – evtl. unter Einbezug der in dieser Arbeit ausgearbeiteten Programmtheorie – anzupassen. Aber auch bei einer Änderung des Angebotsformats oder auch der Programmkonzeption wäre es sinnvoll, die praktischen Impressionen und die daraus rekonstruierte Programmtheorie heranzuziehen und produktiv nutzbar zu machen. Gleichzeitig sollte mit dem Angebotsformat konkret die Rolle der quantitativen Diagnostiktests und deren Integration oder Separation mitgedacht werden. So lässt sich auch vorstellen, Teile des Angebots mehr in Form von „Trainings“ zu konzipieren, die gezielt die programmrelevanten Fähigkeiten trainieren. Dazu ergänzend kann den Kindern komplementär gewisse Freiräume geboten werden, in denen weniger stark Fähigkeitsveränderungen, sondern Darbieten von Freiräumen im Vordergrund stehen. Es lässt sich auch vorstellen, dies abwechselnd zu gestalten, d. h. eine Phase mit einem

konkreten Trainingsangebot und eine andere Phase, in der den Kindern vermehrt Freiräume angeboten werden. Hierdurch könnten auch die unterschiedlichen professionellen Hintergründe des Fachpersonals optimal eingesetzt werden.

Standardisierung:

Die bisherige Argumentation des Autors dieser Studie zusammengefasst und angelehnt an StEG, lässt sich argumentieren, dass eine Standardisierung des Programms von Vorteil erscheint – insbesondere, wenn konkrete Wirkungen in Form von Wirkungslogiken oder ähnlichem beschrieben werden. Als Gegenargument in Bezug auf die Standardisierung und Ausarbeitung eines konkreten Programmangebots könnte eingewandt werden, dass teilweise Schwierigkeiten auftreten können, auf Spontaneität in den sozio-emotionalen und motivationalen Voraussetzungen der Schülerinnen und Schüler reagieren sowie die angestrebte Beziehungsqualität herstellen und aufrechterhalten zu können. Darüber hinaus beinhaltet eine Standardisierung weitere Aspekte wie einen zeitintensiven und gründlichen Vorbereitungsprozess des Programms, welcher möglichst alle zukünftig Tätigen – und möglichst auch alle bisher Tätigen – einbezieht. Gleichzeitig bedarf dies auf anderer Seite einen engeren Austausch mit der jeweiligen Schule, um konkret benötigte räumliche und zeitliche Ressourcen sicherzustellen. Hierbei müsste allerdings noch geklärt werden, auf welche Bereiche sich diese Standardisierung auswirkt, ob sie rein inhaltlich oder auch methodisch-didaktisch sein soll, wie sie die verschiedenen Angebote aufeinander und auf den Schulunterricht beziehen, wie die Angebote aufeinander aufbauen, wie sie sich auf die Lernvoraussetzungen und Bedürfnisse der Schülerinnen und Schüler beziehen können, wie die Ergebnisse der quantitativen Diagnostikinstrumente zur formativen Evaluation herangezogen werden können sowie wie dem Fachpersonal konkrete Freiheiten gewahrt werden können.

Bei Beibehaltung der aktuellen Angebotsform besteht, wie bereits erwähnt, das Risiko, dass das Angebot nicht auf die vom Programm intendierten Ziele ausgerichtet ist, da das Fachpersonal aufgrund der gewährten Freiheiten und persönlicher Zieldimensionen eigene Schwerpunkte setzt, die nicht unmittelbar mit den Programmzielen übereinstimmen müssen. Deshalb empfiehlt es sich, bzgl. zukünftiger Programmkonzeption zu überlegen, ob und inwiefern das Programm eigene Zielvorgaben vorgibt und diese auch „überprüft“ haben will. Sollte ein Schwerpunkt auf die inhaltliche und methodisch-didaktische Freiheit des Fachpersonals bei der Angebotskonzeption gelegt werden, so empfehlen sich entweder allgemeine Programmzieldimensionen, die möglichst viele Inhalte miteinbeziehen oder aber konkrete Steuerungsmechanismen, die sowohl dem Programm als auch dem Fachpersonal Informationen über die aktuelle Zielerreichung geben, um das Angebot stärker auf die Programmziele hin auszurichten. Gleichzeitig ist darauf zu achten, das Fachpersonal nicht zu überfordern, indem zu hohe Anforderungen oder Erwartungen gestellt werden. Um dies zu verhindern, empfiehlt es sich, Fachpersonal in den Konzeptionsprozess stark miteinzubeziehen und insbesondere darauf zu achten, ob das eigene Rollenverständnis mit den angestrebten Änderungen vereinbar ist.

Kooperationsprozesse:

Hinsichtlich der Kooperation ist es vorstellbar, dass insbesondere Kommunikations- und Austauschprozesse zwischen dem pädagogischen Fachpersonal und den schulischen Akteurinnen und Akteuren explizit bei der Programmkonzeption mitgedacht werden sollten (vgl. Jugendinstitut, 2019). Es erscheint empfehlenswert, festzulegen, welche

Informationsinhalte über welche Kommunikationswege zu welchen Zeitpunkten an welche Personen fließen sollen. Dies sollte insbesondere gemeinsam mit den schulischen Akteurinnen und Akteuren abgeklärt und vereinbart und in Zusammenarbeit eine gemeinsame Basis von Kommunikationsregeln festgelegt werden (vgl. Jugendinstitut, 2019). Ausgehend von den Kommunikationsprozessen können darauf aufbauend weitere Kooperationsformate bestimmt werden, die die Rahmenbedingungen für die multiprofessionellen Kooperationsformen legen. Hierbei hat StEG einige Vorschläge entwickelt, wie solche Kooperationsformate aussehen und umgesetzt werden könnten (Jugendinstitut, 2019). StEG beschreibt, dass die Kooperation zwischen dem pädagogischen Fachpersonal und den schulischen Akteurinnen und Akteuren eine wichtige Voraussetzung zur Umsetzung von individueller Förderung ist (Jugendinstitut, 2019). Gleichzeitig ermöglicht die Kooperation mit externen Kooperationspartner das Schaffen von neuen Methoden, um den Ganzttag weiterzuentwickeln und auf die Schülerbedarfe einzugehen (Jugendinstitut, 2019). Dies setzt des Weiteren passende räumliche Strukturen und zeitliche Ressourcen voraus (Jugendinstitut, 2019).

Weiterbildung:

Des Weiteren erscheint es empfehlenswert, darauf zu achten, dass das Fachpersonal über die notwendigen Kompetenzen verfügt, um die angestrebten, angebotsspezifischen Programmziele zu erfüllen. Da das Fachpersonal für das Programm wertvolle biographische und professionelle Erfahrungen und Kenntnisse mitbringt, erscheint es ebenfalls sinnvoll, dies in die Angebotskonzeption auf Programmebene zu integrieren. Auch sollten konkrete Lerngelegenheiten für das Fachpersonal hergestellt werden, die die Personalentwicklung unterstützen, insbesondere in Anbetracht der multiprofessionellen Kooperation. Insbesondere könnte hierbei auch ein Fokus auf die in dieser Arbeit beschriebenen (Lehrer-)Überzeugungen gerichtet werden, welche einerseits noch weiter deskriptiv spezifiziert und andererseits gezielt mit Personalentwicklungsmaßnahmen beeinflusst werden können.

Generierung von Evidenzen zum Programmerfolg:

Zur weiteren Generierung von Daten zur Evaluation des Programmerfolgs sollte insbesondere auf die standardisiertere Erhebung der Diagnostiktests geachtet werden, eine Kontrollgruppe herangezogen werden, das Programm stärker durch eine Programmtheorie fundiert werden und ebenfalls kontinuierlich wissenschaftlich begleitet werden.

8. Fazit

Das Ziel dieser Arbeit war es einerseits, ein summatives Urteil über den Evaluationsgegenstand zu fällen. Die hierfür analysierten Daten ermöglichen die Identifikation verschiedener Indizien, die eine partielle Wirkung des Programms vermuten lassen. Ob eine Wirkung tatsächlich besteht und wie stark diese ausgeprägt ist, ließ sich allerdings mit dem gegebenen Evaluationsdesign nicht beantworten. Für zukünftige Evaluationen erscheint es insbesondere interessant, den Ansatz der theoriebasierten Evaluation heranzuziehen, um Wirkungsaussagen zu approximieren. Dies verlangt allerdings zunächst eine sorgfältig ausformulierte Programmtheorie. Gleichzeitig erweist sich die große Bereitschaft seitens GmS und ENB zur wissenschaftlichen Evaluation als eine gute Möglichkeit zur Etablierung einer Wissenschaft-Praxis-Kooperation, von welcher beide Seiten profitieren können. Gleichzeitig bieten sich solche Gelegenheiten insbesondere für den wissenschaftlichen Nachwuchs an,

erste Erfahrungen mit Praxispartnern zu sammeln und das erworbene methodische und theoretische Wissen in die Praxis zu überführen. Für Stiftungen scheint dies insbesondere dann lohnenswert, wenn keine Geldmittel zur Verfügung stehen, um eine Evaluierung ihres Angebots in Auftrag zu geben.

Literaturverzeichnis:

- Arnoldt, B. (2011). *Kooperation zwischen Ganztagschule und außerschulischen Partnern. Entwicklung der Rahmenbedingungen* (pp. 312-329).
- Ball, S. J. & Junemann, C. (2012). *Networks, new governance and education*. Bristol: Policy Press.
- Baumert, J., Kunter, M. (2011). *Das Kompetenzmodell von COACTIV*. In: Kunter, M., Blumer, H. (1966). Sociological implications of the thought of George Herbert Mead. *American journal of sociology*, 71(5), 535-544.
- Bortz, J., & Schuster, C. (2010). *Statistik für Human-und Sozialwissenschaftler*. Springer-Verlag.
- Brake, A., & Büchner, P. (2012). *Bildung und soziale Ungleichheit: Eine Einführung*. Kohlhammer-Urban-Taschenbücher: Band 700. Stuttgart: Verlag W. Kohlhammer.
- Brecher, C. & Wise, O. (2008). *Looking a gift horse in the mouth. Challenges in managing philanthropic support for public services*. *Public Administration Review*, 68 (6), S146-S161. <https://doi.org/10.1111/j.1540-6210.2008.00985>
- Burkard, C., & Eikenbusch, G. (2005). *Praxishandbuch Evaluation in der Schule* (4. Aufl.). Berlin: Cornelsen Scriptor.
- Clarke, V., & Braun, V. (2013). *Successful qualitative research: A practical guide for beginners*.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences 2nd ed.* Hillsdale, NJ: L. Erlbaum Associates.
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.
- Crosby, B. & Bryson, J. (2010). *Integrative leadership and the creation and maintenance of cross-sector collaborations*. *The Leadership Quarterly* 21(2): 211–230.
- Ditton, H. (2010). *Der Beitrag von Schule und Lehrern zur Reproduktion von Bildungsungleichheit*. In R. Becker & W. Lauterbach (Eds.). *Bildung als Privileg: Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit* (4th ed., pp. 243–272). Wiesbaden: VS Verl. für Sozialwissenschaften.
- Ditton, H. (2013). *Bildungsgerechtigkeit: Analysen zu einer vergessenen Voraussetzung*. In E. Zwick & N. J. Hofmann (Eds.), *Theologie: Vol. 35. Dialog der Religionen: Eine interdisziplinäre Annäherung* (pp. 155–185). Berlin: LIT.
- Dresing, T., & Pehl, T. (2018). *Praxisbuch Interview, Transkription & Analyse. Anleitungen und Regelsysteme für qualitativ Forschende* (8. Aufl.). Marburg: Eigenverlag.
- Drexler, D. (2013). *Qualität im Grundschulunterricht: der Einfluss der Elementar- auf die Primärpädagogik*. Springer-Verlag.
- Etzioni, A. (1973). *The third sector and domestic missions*. *Public Administration Review*, 33 (4), 314. <https://doi.org/10.2307/975110>
- Evers, A. & Laville, J.-L. (Hrsg.). (2004). *The third sector in Europe*. Cheltenham, UK: Edward Elgar.

- Eyal, O. & Yarm, M. (2018). *Schools in cross-sector alliances. What do schools seek in part-nerships?* Educational Administration Quarterly, 54 (4), 648–688. <https://doi.org/10.1177/0013161X18765268>
- Fischer, N., Holtappels, H. G., Stecher, L., & Züchner, I. (2011). *Theoretisch-konzeptionelle Bezüge - ein Analyserahmen für die Entwicklung von Ganztagschulen*. In N. Fischer, H. G. Holtappels, E. Klieme, T. Rauschenbach, L. Stecher, & I. Züchner (Eds.), *Studien zur ganztägigen Bildung. Ganztagschule: Entwicklung, Qualität, Wirkungen: Längsschnittliche Befunde der Studie zur Entwicklung von Ganztagschulen (StEG)* (pp. 18–29). Weinheim und Basel: Beltz Juventa.
- Floerecke, P., & Holtappels, H. G. (2004). *Qualitätsentwicklung in der Kooperation von Jugendhilfe und Schule*. In: Hartnuß, B./Maykus, S.(Hg.): *Handbuch Kooperation von Jugendhilfe und Schule*. Frankfurt/M, 897-922.
- Gängler, H. (2011). *Wozu Ganztagschule?: Zu strukturellen und inhaltlichen Veränderungen im Bildungssystem*. In H. Gängler & T. Markert (Eds.), *Studien zur ganztägigen Bildung. Vision und Alltag der Ganztagschule: Die Ganztagschulbewegung als bildungspolitische Kampagne und regionale Praxis* (315-232). Weinheim, München: Juventa Verlag.
- Gautreaux, M. & Delgado, S. (2016). *Portrait of a Teach for All (TFA) teacher. Media narra-tives of the universal TFA teacher in 12 countries. education policy analysis archives*, 24 (110). <https://doi.org/10.14507/epaa.24.2149>
- Groeben, N. (1986). *Handeln, Tun, Verhalten als Einheiten einer verstehend-erklärenden Psychologie: wissenschaftstheoretischer Überblick und Programmwurf zur Integration von Hermeneutik und Empirismus* (p. 477). Francke.
- Groeben, N., & Scheele, B. (2010). *Das Forschungsprogramm Subjektive Theorien* (pp. 151-165). VS Verlag für Sozialwissenschaften.
- Groeben, N., Wahl, D., Schlee, J., & Scheele, B. (1988). *Das Forschungsprogramm subjektive Theorien: eine Einführung in die Psychologie des reflexiven Subjekts* (p. 364). Francke.
- Guba, E. G., & Lincoln, Y. (Hrsg.). (1989). *Fourth generation evaluation*. London: Sage.
- Haberl, A., Fleiß, J., Kowald, D., & Thalmann, S. (2024). *Take the aTrain. Introducing an interface for the Accessible Transcription of Interviews*. *Journal of Behavioral and Experimental Finance*, 41, 100891.
- Harrow, J. (2011). *Governance and isomorphism in local philanthropy*. *Public Management Review*, 13 (1), 1–20. <https://doi.org/10.1080/14719037.2010.501617>
- Helfferich, C. (2011). *Die Qualität qualitativer Daten* (Vol. 4). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Helmke, A. (2003). *Christoph von Burkard/Gerhard Eikenbusch: Praxishandbuch Evaluation in der Schule*. Berlin: Cornelsen Scriptor 2000. 208 S.[Rezension]. *Zeitschrift für Pädagogik*, 49(1), 148-150.
- Hense, J. U., & Mandl, H. (2003). *Selbstevaluation - Ein Ansatz zur Qualitätsverbesserung pädagogischer Praxis und seine Umsetzung am Beispiel des Modellversuchprogramms SEMIK*. <https://doi.org/10.5282/UBM/EPUB.288>

- Hettlage, R., & Steinlin, M. (2006). *The critical incident technique in knowledge management-related contexts*. Ingenious Peoples Knowledge, 1-18.
- Horvath, A., & Powell, W. W. (2016). *Contributory or disruptive: Do new forms of philanthropy erode democracy?* In R. Reich, C. Cordelli, & L. Bernholz (Eds.), *Philanthropy in democratic societies: History, institutions, values* (pp. 87–124). Chicago, IL: The University of Chicago Press.
- Jugendinstitut, D. (2016). *Ganztagsschule: Bildungsqualität und Wirkungen außerunterrichtlicher Angebote. Ergebnisse der Studie zur Entwicklung von Ganztagsschulen 2012-2015*.
- Jugendinstitut, D. (2019). *Individuelle Förderung: Potenziale der Ganztagsschule*. Frankfurt am Main: Das Konsortium der Studie zur Entwicklung von Ganztagsschulen (StEG).
- Kassambara A (2023). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. R package version 0.7.2, <https://rpkgs.datanovia.com/rstatix/>.
- Kleickmann, T., Vehmeyer, J., Möller, K. (2010). *Zusammenhänge zwischen Lehrervorstellungen*.
- Kolleck, N. (2017). *How (German) foundations shape the concept of education. Towards an understanding of their use of discourses. Discourse: Studies in the Cultural Politics of Education*, 38 (2), 249–261. <https://doi.org/10.1080/01596306.2015.1105789>
- Koranyi, F. & Kolleck, N. (2018). *Roles of local foundations in German community-based initiatives. Devolving criticism of philanthropy to the local?* education policy analysis archives, 26 (139). <https://doi.org/10.14507/epaa.26.3683>
- Koranyi, F. (2021). *Meta-Governance in Bildungslandschaften und rollenbasierte Handlungsfelder lokaler Bildungstiftungen*. Freie Universität Berlin (Germany).
- Kuhn, T (1962). *The structure of scientific revolutions*. Chicago: University press.
- Kultusministerkonferenz, (KMK). (2002). *PISA 2000 – Zentrale Handlungsfelder* (Beschluss der 299. Kultusministerkonferenz vom 17./18.10.2002).
- Kultusministerkonferenz, (KMK). (2018). *Allgemeinbildende Schulen in Ganztagsform in den Ländern in der Bundesrepublik Deutschland: Statistik 2012 bis 2016*. Berlin. Abgerufen von: https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Dokumentationen/TS_2016_Bericht.pdf
- Kultusministerkonferenz, (KMK). (2023). *Allgemeinbildende Schulen in Ganztagsform in den Ländern in der Bundesrepublik Deutschland: Statistik 2017 bis 2021* (S. 87). Berlin: Sekretariat der Kultusministerkonferenz.
- Kultusministerkonferenz, (KMK). (2023). *Empfehlungen zur Weiterentwicklung der pädagogischen Qualität der Ganztagschule und weiterer ganztägiger Bildungs- und Betreuungsangebote für Kinder im Grundschulalter*. (Beschluss der Kultusministerkonferenz vom 12.10.2023). Abgerufen von: https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2023/2023_10_12-Ganztag-Empfehlung.pdf
- Kurz, B., Kubek, D., & Schultze, S. (2013). *Kursbuch Wirkung*.
- Lamnek, S. (2005). *Qualitative Sozialforschung*. 4., vollst. überarb. Aufl. Basel: Beltz.

- Lenhard, W., Lenhard, A. & Schneider, W. (2018). *ELFE II. Ein Leseverständnistest für Erst- bis Siebtklässler - Version II*. Hogrefe Verlag.
- Lewalter, D., Diedrich, J., Goldhammer, F., Köller, O., & Reiss, K. (Eds.). (2023). *PISA 2022: Analyse der Bildungsergebnisse in Deutschland*. Waxmann Verlag.
- Linberg, T., Struck, O., & Bäumer, T. (2018). Vorzug Ganztagschule? Zusammenhänge mit der Kompetenzentwicklung im Bereich Lesen und Mathematik. *Zeitschrift für Erziehungswissenschaft*, 21(6), 1205-1227.
- Mandl, H., & Hense, J. (2007). *Lässt sich Unterricht durch Evaluation verbessern?* In W. Schönig (Ed.), *Spuren der Schulevaluation: Zur Bedeutung und Wirksamkeit von Evaluationskonzepten im Schulalltag* (pp. 85–99). Bad Heilbrunn: Klinkhardt.
- Mayring, P. & Fenzl, T. (2024). *QCAmap // a software for Qualitative Content Analysis*. <https://www.qcmap.org/ui/de/home>
- Mayring, P. (2022). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (13. Neuausgabe). Beltz.
- Meyer, H.-D. & Boyd, W. L. (2001). *Civil society, pluralism, and education. Introduction and overview*. In H.-D. Meyer & W. L. Boyd (Hrsg.), *Education between state, markets, and civil society. Comparative perspectives* (S. 1–13). New York, NY: Routledge, Taylor & Francis Group.
- Neubrand, M. (Hrsg.). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des*
- Nixdorf, C. P. (2020). *Handlungskompetenz erfassen mit der Critical Incident Technique*.
- Pajares, F. (1992). *Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct*. In: *Review of Educational Research*, 62 (3), S. 307-332.
- Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.
- R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.3.0) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rauschenbach, T. (Ed.). (2012). *Ganztagschule als Hoffnungsträger für die Zukunft?: Ein Reformprojekt auf dem Prüfstand; Expertise des Deutschen Jugendinstituts (DJI)*. Gütersloh: Verl. Bertelsmann Stiftung.
- Reuschenbach, B. (2008). *Einfluss von Expertise auf Problemlösen und Planen im komplexen Handlungsfeld Pflege*. Logos-Verlag.
- Salamon, L. M. & Anheier, H. K. (1997). *Defining the nonprofit sector. A cross-national analysis*. Manchester: Manchester University Press.
- Schaller, F. (2020). *Die Politik der Bildungsarmut: Schul- und Bildungspolitik in Deutschland seit PISA*. Nomos Verlagsgesellschaft mbH & Co. KG. <https://doi.org/10.5771/9783748905097>
- Schulz-Gade, H. (2015). *Ganztagsschulbildung in Deutschland – ausgewählte Aspekte zur Entwicklung und aktuellen Situation*. *Bildung und Erziehung*, 68(2), 237–254. <https://doi.org/10.7788/bue-2015-0208>

- Spinath, B., Stiensmeier-Pelster, J., Schöne, C. & Dickhäuser, O. (2012). *SELLMO. Skalen zur Erfassung der Lern- und Leistungsmotivation. 2., überarbeitete und neu normierte Auflage*. Hogrefe Verlag
- Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S., & Henschel, S. (2022). *Zusammenfassung und Einordnung der Befunde*. In P. Stanat, S. Schipolowski, R. Schneider, K. A. Sachse, S. Weirich, & S. Henschel (Hrsg.), *IQB-Bildungstrend 2021: Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich* (S. 259–284). Münster; New York: Waxmann. <https://doi.org/10.31244/9783830996064>
- Stockmann, R. & Meyer, M. (2014). *Evaluation. Eine Einführung (2. Aufl.)*. Opladen/Toronto: Barbara Budrich.
- Stockmann, R. (2006). *Evaluation und Qualitätsentwicklung: eine Grundlage für wirkungsorientiertes Qualitätsmanagement*. Waxmann Verlag.
- Straub, J., & Weidemann, D. (2015). *Handelnde Subjekte. „Subjektive Theorien“ als Gegenstand der verstehend-erklärenden Psychologie*. Gießen: Psychosozial-Verlag.
- Strietholt, R., Manitius, V., Berkemeyer, N., & Bos, W. (2015). *Bildung und Bildungsungleichheit an Halb- und Ganztagschulen*. *Zeitschrift für Erziehungswissenschaft*, 18(4), 737–761. <https://doi.org/10.1007/s11618-015-0634-6>
- Strübing, J. (2018). *Qualitative Sozialforschung: Eine komprimierte Einführung*. Berlin, Boston: De Gruyter.
- Terhart, E. (1978). *Interpretative Unterrichtsforschung: kritische Rekonstruktion und Analyse konkurrierender Forschungsprogramme der Unterrichtswissenschaft*. Klett-Cotta.
- Thielen, M. (2010). *Soziale Arbeit im Kontext Schule–Möglichkeiten und Grenzen der Kooperation von Schul- und Sozialpädagogik in praxisbezogenen Abgangsklassen unterer Bildungsgänge*. *Zeitschrift für Sozialpädagogik*, 8(1), 7-26.
- Thümler, E., Bögelein, N. & Beller, A. (2014). *Education philanthropy in Germany and the United States*. In E. Thümler, N. Bögelein, A. Beller & H. K. Anheier (Hrsg.), *Philanthropy and education. Strategies for impact* (S. 3–28). Basingstoke: Palgrave Macmillan.
- Tillmann, K., Sauerwein, M., Hannemann, J., Decristan, J., Lossen, K., & Holtappels, H. G. (2018). *Förderung der Lesekompetenz durch Teilnahme an Ganztagsangeboten? – Ergebnisse der Studie zur Entwicklung von Ganztagschulen (StEG)*. In M. Schüpbach, L. Frei, & W. Nieuwenboom (Hrsg.), *Tagesschulen* (S. 289–307). Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-19024-8>
- Van Santen, E., & Seckinger, M. (2003). *Kooperation in der Kinder- und Jugendhilfe: zwischen Anspruch und Wirklichkeit: Eine qualitative Feldstudie*. *Qualitative Forschung in der Sozialpädagogik*, 119-144.
- Vogel, P. (2019). *Grundbegriffe der Erziehungs- und Bildungswissenschaft*. Opladen & Toronto: Verlag Barbara Budrich.
- Von Kardorff, E., & Schönberger, C. (2020). *Qualitative Evaluationsforschung* (pp. 135-157). Springer Fachmedien Wiesbaden.

- Voss, T., Kleickmann, T., Kunter, M., Hachfeld, A. (2011). *Überzeugungen von Mathematiklehrkräften*.
- Wagner, R. F. (2016). *Unterricht aus Sicht der Lehrerinnen und Lehrer: Subjektive Theorien zur Unterrichtsgestaltung und ihre Veränderung durch ein Training zu neuen Unterrichtsmethoden*. Julius Klinkhardt.
- Weiss, C. H. (1995). *Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families*. *New approaches to evaluating community initiatives: Concepts, methods, and contexts*, 1, 65-92.
- Wiere, A. (2011). *Wie wirkt die Ganztagschule?: Forschungsfragen und Befunde*. In H. Gängler & T. Markert (Eds.), *Studien zur ganztägigen Bildung. Vision und Alltag der Ganztagschule: Die Ganztagschulbewegung als bildungspolitische Kampagne und regionale Praxis* (pp. 33–58). Weinheim, München: Juventa Verlag.
- Winter, R. (2010). *Symbolischer Interaktionismus*. *Handbuch qualitative Forschung in der Psychologie*, 79-93.
- Wottawa, H., & Thierau-Brunner, H. (2003). *Lehrbuch Evaluation (3. Aufl.)*. Bern: Huber.
- Yemini, M. & Sagie, N. (2015). *School-nongovernmental organization engagement as an entrepreneurial venture. A case study of Sunlight's engagement with Israeli schools*. *Educational Administration Quarterly*, 51 (4), 543–571.
- Yemini, M., Cegla, A. & Sagie, N. (2018). *A comparative case-study of school-LEA-NGO inter-actions across different socio-economic strata in Israel*. *Journal of Education Policy* 33(2): 243–261.
- Züchner, I., & Fischer, N. (2011). *Ganztagschulentwicklung und Ganztagschulforschung: Eine Einleitung*. In N. Fischer, H. G. Holtappels, E. Klieme, T. Rauschenbach, L. Stecher, & I. Züchner (Eds.), *Studien zur ganztägigen Bildung. Ganztagschule: Entwicklung, Qualität, Wirkungen: Längsschnittliche Befunde der Studie zur Entwicklung von Ganztagschulen (StEG)* (pp. 9–17). Weinheim und Basel: Beltz Juventa.

Anhang

Zu erhebende Konstrukte:

Allgemeines Studiendesign:

- Mixed-Method

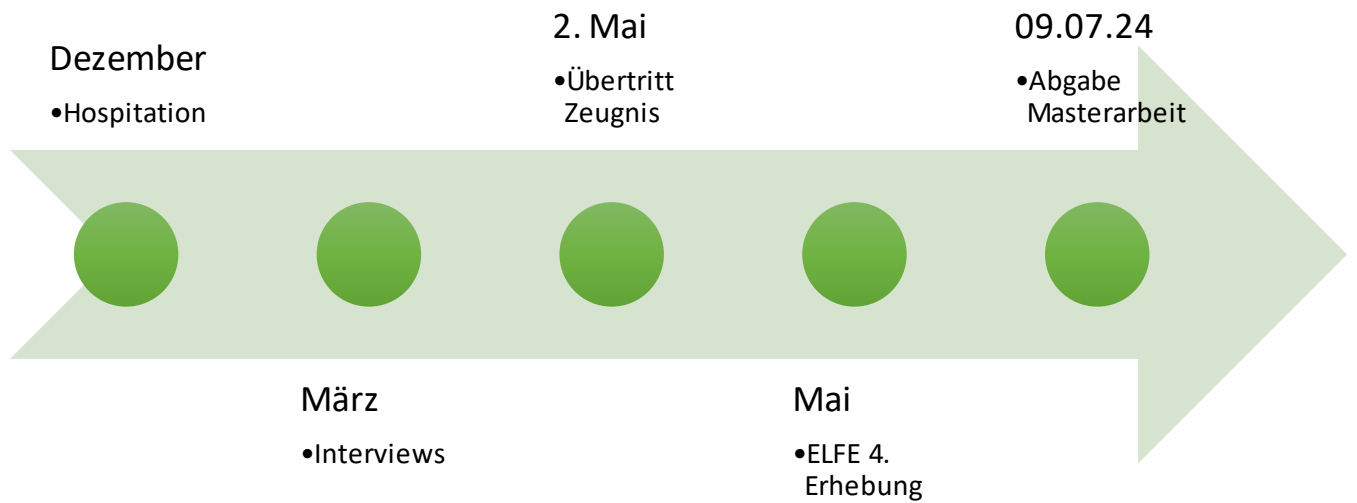
Quantitative Analyse:

- Zur Verfügung stehende Daten:
 - *ELFE II*: 1.-4. Schuljahr
 - *SELLMO*: 3.-4. Schuljahr
 - *Noten*: 2.-4. Schuljahr
 - *BEO*: 2.-4. Schuljahr
 - *Übertrittsempfehlungen*: Mai
- Fragestellungen nach thematischen Bereichen:
 - 1. Chancengleichheit erhöhen und Abbau von Bildungsbenachteiligung:**
 - Hängt der Lernerfolg von sozialen Variablen ab?
 - Sind die Lernfortschritte für alle gleich groß?
 - Wird ein Mindestniveau in den Lesekompetenzen erreicht?
 - 2. Wirkungsorientierung des Programms:**
 - Weisen die SuS eine signifikante Verbesserung in den Zielvariablen auf?
 - Lässt sich eine Verbesserung in den Noten erkennen?
 - 3. Talentförderung & individuelle Förderung:**
 - Profitieren SuS mit unterdurchschnittlichen Leseleistungen von der (individuellen) Förderung?
 - 4. Übergang auf SK I:**
 - Erreichen die SuS die erforderlichen Kompetenzen für den Übergang auf die Realschule und das Gymnasium?

Qualitative Analyse:

- Zu interviewende Personen: 4 Fachpersonal, 1 Klassenlehrkraft
- Interviewdauer: 30-40 min
- Fragestellungen:
 1. Wie geht das Fachpersonal mit **quantitativen Daten** zu Lernverläufen um?
 2. Wie werden **Qualitätssicherungsprozesse** realisiert?
 3. Wie nimmt das Fachpersonal den **multiprofessionellen Umgang** mit den anderen schulischen Akteuren wahr?
 4. Wie schätzt das Fachpersonal den **Nutzen von aktuellen Fördermaßnahmen** für zukünftige Programme ein?
 - Im Vergleich zum Nutzen des Regelunterrichts
 - Einschätzung bzgl. Hochskalierung der Fördermaßnahmen oder Individualisierung ausbauen

Timeline:



Schülerprofil:

Schülerprofil

Schülerdaten:

Name:	106
Geschlecht:	Weiblich
Gesprochene Sprache zuhause:	Gemischt
Wiederholung 1. Klasse:	Nein
Einzelförderung:	wöchentlich

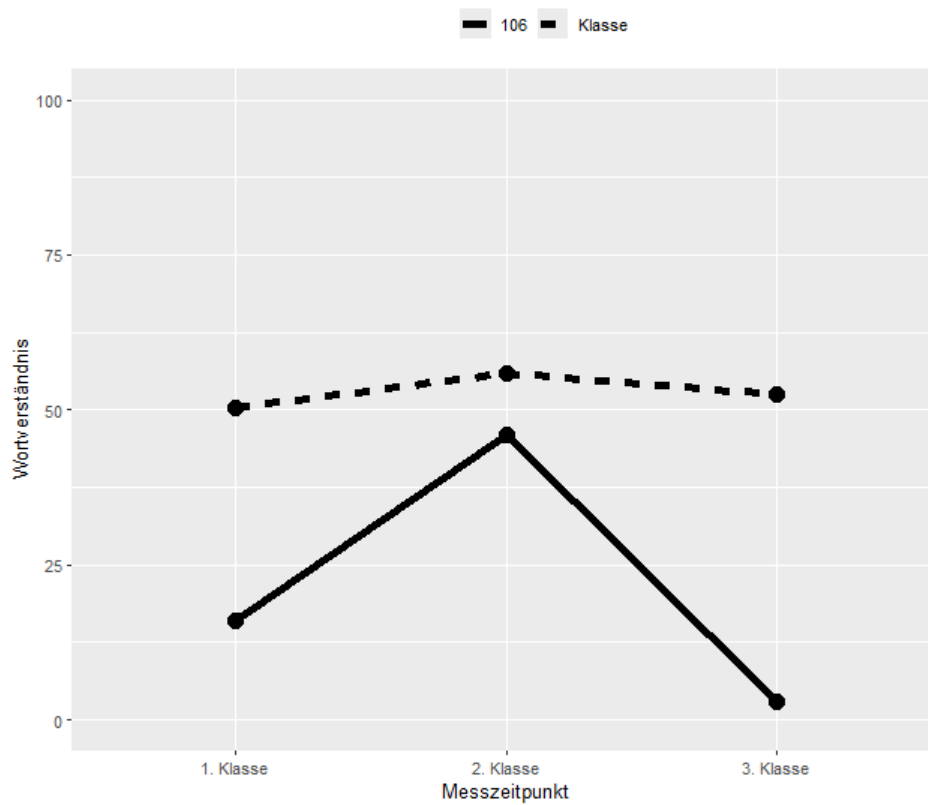
Deutschnote:

	106	Klasse
1. Klasse	4	2,5
2. Klasse	4	2,7
3. Klasse	5	2,7

Lernmotivation*:

	106	Klasse
3. Klasse	32,6	43,3

Wortverständnis*:



Einverständniserklärung qualitative Interviews:



Einverständniserklärung zum Interview

Forschungsprojekt: Evaluation des Projekts „Eine Note besser“
Durchführende Institution: LMU München
Projektleitung/Datenschutz: Fabian Feselmayer
Interviewer: Fabian Feselmayer
Interviewdatum:

Ich erkläre mich dazu bereit, im Rahmen des genannten Forschungsprojekts an einem Interview teilzunehmen. Ich wurde über das Ziel und den Verlauf des Forschungsprojekts informiert.

Ich bin damit einverstanden, dass das Interview mit einem Aufnahmegerät aufgezeichnet und durch ein Schreibbüro in Schriftform gebracht wird. Die Audiodateien werden unter einem Token gespeichert und zum Projektende am 08.07.2024 gelöscht. Die Transkripte der Interviews werden anonymisiert, d.h. ohne Namen und Personenangaben gespeichert. Die wissenschaftliche Auswertung des Interviewtextes erfolgt durch die Projektleitung. Die LMU München arbeitet nach den Vorschriften der Datenschutz-Grundverordnung (DS-GVO), des Bundesdatenschutzgesetzes (BDSG (neu) 2018) vom 25.05.2018 sowie allen anderen datenschutzrechtlichen Bestimmungen.

Ich bin damit einverstanden, dass einzelne Sätze aus den Transkripten, die nicht mit meiner Person in Verbindung gebracht werden können, als Material für wissenschaftliche und unterrichtende Zwecke genutzt werden können. Des Weiteren stimme ich zu, dass nach Projektende in darauf aufbauenden Forschungsprojekten die anonymisierten Transkripte verwendet werden können.

Meine Teilnahme an der Erhebung und meine Zustimmung zur Verwendung der Daten, wie oben beschrieben, sind freiwillig. Ich habe jederzeit die Möglichkeit, meine Zustimmung zu widerrufen. Durch Verweigerung oder Widerruf entstehen mir keine Nachteile. Ich habe das Recht auf Auskunft, Berichtigung, Sperrung und Löschung, Einschränkung der Verarbeitung, Widerspruch gegen die weitere Verarbeitung sowie auf Datenübertragbarkeit meiner personenbezogenen Daten.

Unter diesen Bedingungen erkläre ich mich bereit, das Interview zu geben, und bin damit einverstanden, dass es aufgezeichnet, verschriftlicht, anonymisiert und ausgewertet wird.

Ort, Datum, Unterschrift Interviewte_r

Ort, Datum, Unterschrift Interviewer_in

Leitfaden für das Fachpersonal:

1. Einleitung

1. Einführendes Gespräch & eigene Vorstellung
2. Erläutern des Interviewablaufs
 - ➔ Anonymisiert, man kann keine Rückschlüsse auf die Person ziehen
 - ➔ Ziel ist es, Herauszufinden, welche Dinge gut funktionieren, welche Dinge noch verbessert werden können
 - ➔ Es gibt keine richtigen oder falschen Antworten, gerne mehr Details als weniger Details angeben.
 - ➔ Ich stelle die Fragen absichtlich sehr offen, weil es mir wichtig ist, deinen Blickwinkel auf das Thema zu verstehen.
3. Einverständniserklärung
4. Starten der Aufnahme

2. Unterrichtskonzeption

- Könntest du dich zunächst noch einmal kurz vorstellen und deine Rolle im Projekt?
 - Rolle
 - Aufgaben und Pflichten
 - Ziele
- Wie findet denn so ein typischer Entscheidungsprozess statt, was du in einem Unterricht machst?
 - Konkrete methodisch-didaktische Entscheidungen
 - Vorbereitung & Nachbereitung
 - *Fördermaßnahmen für einzelne Schüler*
 - Programm: Freiheit bei der Methodenwahl
 - Lernziele
 - Absprachen mit anderen Fachkräften
 - Unterstützungsangebote seitens des Programms
- Würdest du dir wünschen, mehr Freiheiten/Vorgaben bei der Wahl der Methoden zu haben?
 - Mitbestimmung bei Methodenwahl

3. Qualität

- Jetzt eine eher abstrakte Frage: Was ist denn für dich Bildungsqualität?
 - Konkreter Unterrichtskontext
 - GMS-Qualitätsverständnis
- Wie würdest du pädagogische Handlungen beschreiben, die im Unterrichtskontext wirksam sind?
 - Bezug auf Qualität
 - Erkenntnis über Wirksamkeit der eigenen Handlungen
 - Evidenzbasierung

- Diagnostik und Evaluation

4. Evidenzbasierung

- Wie sinnvoll oder umsetzbar findest du es, wissenschaftliche Erkenntnisse in den Unterrichtsalltag zu integrieren?
 - Neue wissenschaftliche Erkenntnisse
 - Wissenschaftliche Instrumente zur Diagnose und Evaluation
 - Umsetzbarkeit
 - Schwachpunkte von wissenschaftlichen Evidenzen
 - Eigene Entscheidungsfindung: wie evidenzbasiert?
- Wie würdest du denn die jetzige Klasse leistungsmäßig beschreiben?
 - Einzelne Schüler, die herausstechen?
 - Sprachliche Entwicklung
 - Motivation
- Im Rahmen von „Eine Note besser“ wurden bestimmte Leistungstests eingesetzt, um Daten zur Leistungsentwicklung der Kinder zu sammeln. Ein Ziel könnte es sein, die Daten bei Entscheidungen für Fördermaßnahmen im Unterrichtsalltag zu verwenden. Hierfür haben wir einen möglichen Ansatz erstellt, den wir euch Dir gerne vorstellen wollen. Ich würde ihn dir zunächst erläutern, dann würde ich dich fragen, wie du die dargestellten Ergebnisse interpretieren würdest und am Schluss kannst du uns gerne noch Feedback über den Vorschlag geben.
 - Eigene Kompetenz im Umgang mit Test (Was bräuchte man?)
 - Andere Ideen, wie evidenzbasierte Instrumente in den Alltag integriert werden könnten?

5. Programm

- Programm „Eine Note besser“: Was hältst du so allgemein vom Programm?
 - Wirksamkeit
 - Gut funktionierende Dinge
 - Verbesserungswürdige Dinge
 - Passung mit dem spezifischen Schulkontext
 - Belastungserleben durch strukturelle Bedingungen
 - Vorgaben hinsichtlich Qualität und Realisierbarkeit
 - Einbettung in Schulalltag
 - Rollenkonflikte
- Inwiefern denkst du, dass sich das Programm auf andere Schulen übertragen ließe?
 - Was können andere Projekte von „Eine Note besser“ lernen?
- Multiprofessionelle Kooperation: Wie empfindest du die Zusammenarbeit mit den Mitarbeitenden des Programms?
 - Absprachen
 - Lehrkraft
- Gibt es noch Themen, die dir wichtig sind und über die wir noch nicht gesprochen haben?

Leitfaden für die Lehrkraft:

1. Einleitung

5. Einführendes Gespräch & eigene Vorstellung
6. Erläutern des Interviewablaufs
 - ➔ Anonymisiert, man kann keine Rückschlüsse auf die Person ziehen
 - ➔ Ziel ist es, Herauszufinden, welche Dinge gut funktionieren, welche Dinge noch verbessert werden können
 - ➔ Es gibt keine richtigen oder falschen Antworten, gerne mehr Details als weniger Details angeben.
 - ➔ Ich stelle die Fragen absichtlich sehr offen, weil es mir wichtig ist, deinen Blickwinkel auf das Thema zu verstehen.
7. Einverständniserklärung
8. Starten der Aufnahme

2. Programm

- Programm „Eine Note besser“: Was hältst du so allgemein vom Programm?
 - Wirksamkeit
 - Gut funktionierende Dinge
 - Verbesserungswürdige Dinge
 - Vorgaben hinsichtlich Qualität und Realisierbarkeit
 - Entlastung / Welche Aufgaben übernimmt Programm?
 - Passung mit Schulkontext
- Multiprofessionelle Kooperation: Wie empfindest du die Zusammenarbeit mit den Mitarbeitenden des Programms?
 - Absprachen
 - Lehrkraft
 - Konkretes Kooperationsbeispiel

3. Qualität

- Jetzt eine eher abstrakte Frage: Was ist denn für dich Bildungsqualität für ein externes Programm wie „Eine Note besser“?
 - Konkreter Unterrichtskontext
 - Programm: Qualitätsverständnis
- Wie würdest du pädagogische Handlungen beschreiben, die im Unterrichtskontext wirksam sind?
 - Bezug auf Qualität

- Erkenntnis über Wirksamkeit der eigenen Handlungen
 - Evidenzbasierung
 - Diagnostik und Evaluation
- Wie würdest du das Qualitätsverständnis von GMS bzw. „Eine Note besser“ beschreiben?

4. Evidenzbasierung

- Wie sinnvoll oder umsetzbar findest du es, dass externe Programme wie „Eine Note besser“ wissenschaftliche Erkenntnisse in den Unterrichtsalltag integrieren?
- Neue wissenschaftliche Erkenntnisse
 - Wissenschaftliche Instrumente zur Diagnose und Evaluation
 - Umsetzbarkeit
 - Schwachpunkte von wissenschaftlichen Evidenzen
 - Eigene Entscheidungsfindung: wie evidenzbasiert?

Erweiterter Leitfaden für die Programmleitung

Schule

- Einbettung in Schule:
- Welche Ressourcen & Räume?
 - Teamtreffen mit allen schulischen Beteiligten?
 - Einbettung in Kommunikationsabläufe der Schule?
 - Einbettung in schulische Gremien?
 - Zusammenarbeit mit schulpsychologischer Beratung? (Stephanie Kunde)
 - Zusammenarbeit mit Mobiler Sonderpädagogischer Dienst? (Frau Zimmerlein)
 - Schule will Leitbild umsetzen, inwiefern würdet ihr eingebunden?
 - Wie stark würdest du die Kooperation mit der Schule beschreiben? 0-10
 - Wird die Kooperation als gleichberechtigt wahrgenommen?
 - Wie würdest du dir die Zusammenarbeit mit der Schule vorstellen?
- Informationen über Schule:
- Übertrittsquoten
 - Weitere Programme an der Schule (Vergangenheit & Gegenwart)
 - Schulpsychologische und Erziehungsberatung (Bernd Kirchhoff)

Programm

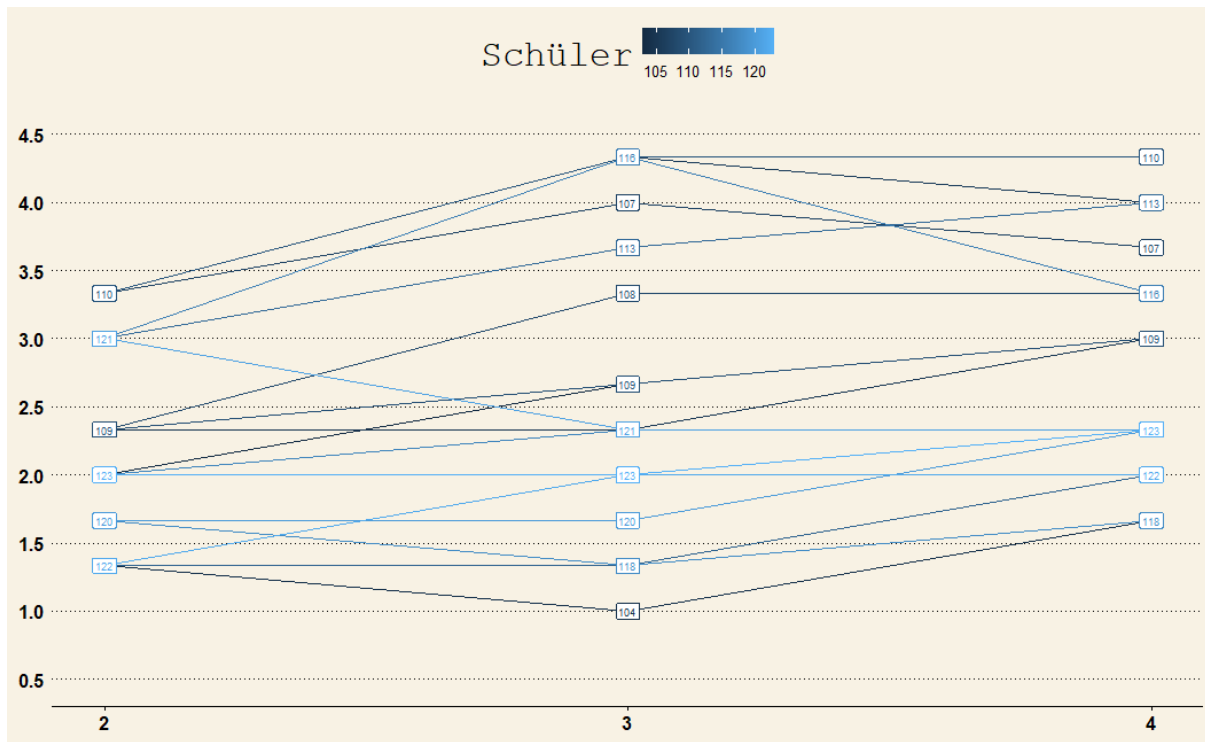
- Programmziel: Chancengerechtigkeit
- Wie definiert ihr Chancengerechtigkeit?
- Qualitätsmanagement:

- Welche Qualitätssicherungs- und -weiterentwicklungsprozesse?
 - Inwiefern wurde das Programm in den letzten vier Jahren angepasst?
 - Wie würdest du die Evidenzbasierung des Programms (Programm beruht auf wissenschaftlichen Standards) beschreiben?
- Maßnahmen der Professionalisierung des Personals:
 - Fort- und Weiterbildungen (mit Lehrkräften)
 - Inwiefern hat Personal die Kompetenz, Grundschul Kinder mit schwierigen Startbedingungen zu fördern?
- Kontakt mit Eltern:
 - Angebote & Frequenz

GMS

- Finanzgeber von GMS bzw. ENB:
 - Wie wird ENB finanziert?
 - Wie wird GMS finanziert?
 - Castringius Stiftung: Welche Rolle?
- Struktur von GMS:
 - Unterstützungsangebote
 - Kooperationsvereinbarung mit Schule?
 - Programmkonzeption: Wie fand der Aushandlungsprozess mit der Schule statt?

Individuelle Lernverläufe – Jahresdurchschnittsnote:



Durchschnittliche Gesamtjahresendnoten über alle Messzeitpunkte