

EXTERNE EVALUATION DES PROJEKTS „*RESPECT* *U*“

Eine Kooperation zwischen der Stiftung Gesellschaft
macht Schule und Anna Gieschen, wiss. Mitarbeiterin
der LMU München – Stand Oktober 2021
(Ergebnisbericht)

Anna Gieschen, M.A.

Lehrstuhl für Allgemeine Pädagogik, Erziehungs- und
Sozialisationsforschung

Ludwig-Maximilians-Universität München Leopoldstr.

13 | 80802 München

E-Mail: a.gieschen@edu.lmu.de

Inhaltsverzeichnis

1 Vorwort	3
2 Evaluation des Projekts „respect U“	3
2.1 Evaluationsgegenstand und -bereich	3
2.2 Fragestellungen und Ziele der Evaluation	4
2.3 Ablauf der Evaluation	4
2.4 Qualitätsindikatoren	5
2.5 Evaluationsstandards	5
3 Methodik	6
3.1 Design	6
3.1 Stichprobe	7
3.2 Erhebungsinstrumente	7
3.3 Statistische Auswertung	9
4 Ergebnisdarstellung und -interpretation	11
4.1 Messzeitpunkt 1	11
4.1.1 Deskriptive Statistiken	12
4.1.2 Korrelationen	12
4.1.4 Zusammenfassung Messzeitpunkt 1	15
4.2 Messzeitpunkt 2	15
4.2.1 Korrelationen	16
4.2.2 T-Test für abhängige Stichproben	16
4.2.3 Multiple Regressionen	18
4.2.4 Zusammenfassung Messzeitpunkt 2	20
5 Limitationen	20
5.1 Stichprobe	20
5.2 Methodik	21
5.3 Auswertung	21
6 Fazit und Ausblick	23
7 Quellen	24
8 Anhang	25

1 Vorwort

Die Stiftung Gesellschaft macht Schule und der Lehrstuhl für Allgemeine Pädagogik, Erziehungs- und Sozialisationsforschung an der Ludwig-Maximilians-Universität München, haben eine Kooperation vereinbart, in deren Rahmen das Projekt „*respect U*“ evaluiert werden soll.

Zu Beginn der Kooperation wurde die Projektevaluation von einer Studentin der LMU München begleitet, welche im Zuge ihrer Masterarbeit zielgruppenadäquate Erhebungsinstrumente entwickelte, eine Bestandsaufnahme zu Beginn des Schuljahres durchführte sowie Möglichkeiten der zukünftigen Datenerhebung und -analyse erarbeitete. Die Studentin setzte in ihrer anschließenden Tätigkeit als wissenschaftliche Mitarbeiterin das begonnene Evaluationsvorhaben fort und verankerte die zuvor erarbeiteten Analysemöglichkeiten in der weiteren Projektevaluation.

Der vorliegende Evaluationsbericht basiert sowohl auf den Erkenntnissen der Masterarbeit als auch auf der Umsetzung deren Implikationen. Im Folgenden werden der Evaluationsauftrag und die angewendeten Methoden knapp vorgestellt, die Ergebnisse präsentiert und interpretiert sowie Limitationen der Evaluation benannt. Der Fokus liegt hierbei auf der Ergebnispräsentation und -interpretation, wohingegen der vorangegangene Evaluationsbericht (November 2020) primär die theoretische und methodische Dimension des Vorhabens thematisiert. Der nachfolgende Bericht (Juni 2022) wird sich dann mit der Diskussion bzw. Reflexion des Evaluationsprojekts beschäftigen.

2 Evaluation des Projekts „*respect U*“

Zu Beginn wird das externe Evaluationsvorhaben anhand des Gegenstands, der Frage- und Zielstellungen, des Ablaufs, der Qualitätsindikatoren und der Standards der Evaluation beschrieben.

2.1 Evaluationsgegenstand und -bereich

Das Förderprogramm „*respect U*“ der Stiftung Gesellschaft macht Schule bildet den Gegenstand der externen Evaluation. Der Evaluationsgegenstand bezeichnet das, was evaluiert wird (Mandl & Hense, 2007). Das Förderprogramm zielt darauf ab, das Sozial- und Lernverhalten von Kindern und Jugendlichen zu verbessern. Die Teilnehmenden sollen mit Hilfe des Programms Konflikte ohne Aggressionen lösen und grundlegende Werte des Miteinanders verinnerlichen. Aktuell werden „*respect U*“-Kurse an der Grundschule an der St.-Martin-Straße und an der Mittelschule an der Ichostraße angeboten. Das Evaluationsvorhaben fokussiert sich auf die Teamtrainings von „*respect U*“ (“*Team und Spiel*” / “*Team und Sport*” / “*Team und Hund*”) in den Jahrgangsstufen 5 - 8 an der Mittelschule an der Ichostraße, welche den inhaltlichen Kern des Projekts darstellen.

2.2 Fragestellungen und Ziele der Evaluation

Die Fragestellungen der Evaluation wurden aus dem 2. und 3. Wirkungsziel des Projekts abgeleitet und lauten wie folgt:

1. In welchem Ausmaß verbessern die teilnehmenden Jugendlichen ihre personalen und sozialen Kompetenzen?
2. In welchem Umfang achten die Jugendlichen auf ihre eigenen Bedürfnisse und handeln sozial kompetent?

Ziel der externen Evaluation ist es, diese Fragestellungen zu beantworten und einen Beitrag zur Entwicklungs-, Kontroll- und Lernfunktion von Evaluationen nach Hense und Mandl (2003) zu leisten. Im ersten Schritt wurde deshalb im Gespräch mit dem Auftraggeber festgelegt, dass die Autorin im Rahmen ihrer Masterarbeit 1) Erhebungsinstrumente für die Evaluation des Projekts konzipiert, 2) zu Beginn des Schuljahres eine Status-Quo-Analyse der Kompetenzausprägungen durchführt, 3) Einflussfaktoren auf diese - und damit einhergehende Förderpotentiale - identifiziert sowie 4) Möglichkeiten der zukünftigen Datenerhebung und -analyse erarbeitet. Auf diese Art und Weise sollten Implikationen zur Optimierung des Projekts erarbeitet werden (Entwicklungsfunktion). Im zweiten Schritt sollte dann kontrolliert werden, ob die aufgewendeten Mittel ihrem Aufwand gerecht werden (Kontrollfunktion) und Erkenntnisse geschaffen werden, die zur Qualitätssicherung sowie zur Weiterentwicklung des Projekts dienen (Lernfunktion). Deshalb wurde am Ende des Schuljahres 5) mittels eines Vorher-Nachher Vergleichs das Erreichen der Projektziele überprüft und somit die Evaluationsfragestellungen beantwortet. Zudem wurde eine Implikation der Masterarbeit umgesetzt, die Analyse programmspezifischer Einflussfaktoren. Auf Basis dieser und der Reflexion der Evaluation werden dann im weiteren Verlauf der Kooperation 6) Implikationen für das Programm bzw. dessen weitere Evaluation abgeleitet und die Lernfunktion erfüllt.

2.3 Ablauf der Evaluation

Zu Beginn der Kooperation wurden Schüler*innen- und Kursleiter*innen-Fragebögen für die Evaluation des Projekts entwickelt. Im Juli 2020 wurde dann eine Piloterhebung durchgeführt, in welcher die Schüler*innen-Fragebögen getestet wurden. Im Oktober 2020, zu Beginn des Schuljahres, erfolgte dann die erste Datenerhebung. Abschließend wurden am Ende des Schuljahres erneut Daten erhoben.

2.4 Qualitätsindikatoren

Anhand von Qualitätsindikatoren sollte die Beantwortung der Evaluationsfragestellungen (s. oben) erfolgen und das Erreichen der übergeordneten Projektziele bewertet werden. Unter Qualitätsindikatoren versteht man dabei Größen, die die nicht direkt messbare Qualität einer Einheit durch Zahlen (-verhältnisse) abbilden (Burkhard & Eikenbusch, 2005).

Es wurden folgende Qualitätsindikatoren definiert:

Befragte	Perspektive	Indikator	Kompetenzdimensionen	Ziel
Schüler*innen	Selbsteinschätzung	Personale Kompetenz	Selbstwirksamkeit Selbstwahrnehmung Selbstkontrolle	Verbesserung der personalen und sozialen Kompetenzen & Anerkennung der eigenen Bedürfnisse und sozial kompetentes Handeln
		Soziale Kompetenz	Kooperationsfähigkeit Empathiefähigkeit Konfliktfähigkeit	
Kursleiter*innen	Fremdeinschätzung	Personale Kompetenz	Selbstwirksamkeit Selbstwahrnehmung Selbstkontrolle	
		Soziale Kompetenz	Kooperationsfähigkeit Empathiefähigkeit Konfliktfähigkeit	

Die übergeordneten Projektziele können dann als erreicht angesehen werden, wenn die Teilnehmenden nach einem Schuljahr eine positive Tendenz in den Bereichen der personalen und sozialen Kompetenzen zeigen. Diese positive Tendenz wird zum einen anhand der Selbstauskunft der Teilnehmer*innen ersichtlich, zum anderen von den Kursleiter*innen beobachtet.

2.5 Evaluationsstandards

Die externe Evaluation orientiert sich, neben Erkenntnissen aus der Literatur, dem Evaluationsgegenstand, den Fragestellungen und den Qualitätsindikatoren, an den folgenden Evaluationsstandards (Bauer, 2007, S.36):

1. Nützlichkeit

Die Nützlichkeit der Evaluation kann mit ihrem Ziel, der Qualitätssicherung des Projekts, begründet werden.

2. Durchführbarkeit

Für die Erfüllung des Standards der Durchführbarkeit wurde bei der Wahl der Erhebungszeitpunkte darauf geachtet, dass diese den Schulalltag nicht unterbrechen, sondern auf die Zeiten der „*respect U*“-Trainings fielen. Auf diese Art und Weise sollte zusätzlicher Organisationsaufwand für die Schüler*innen und die Kursleiter*innen vermieden werden.

3. Ethische Korrektheit

Die ethischen Standards sollten durch die Gewährleistung des Datenschutzes erfüllt werden. Die Daten wurden anonymisiert erhoben, d.h. jedem*r Teilnehmenden wurde vorab ein anonymer Code zugeteilt. Darüber hinaus wurde auf einen wertschätzenden Umgang mit allen Beteiligten geachtet.

4. Genauigkeit

Die Genauigkeit der Evaluation sollte mit Hilfe der Verwendung von wissenschaftlichen Methoden (s. Kapitel 3) und der Einhaltung ihrer Gütekriterien, der Objektivität, der Reliabilität und der Validität, gesichert werden.

3 Methodik

Im Folgenden wird die Methodik der externen Evaluation entlang deren Design, Stichprobe, Erhebungsinstrumenten und Auswertungsvorhaben vorgestellt.

3.1 Design

Die Ergebnisse der externen Evaluation stützen sich auf folgende Datengrundlage:

Messzeitpunkt 1	Messzeitpunkt 2
Programminterne Daten	
Schüler*innen Fragebogen (Selbsteinschätzung)	Schüler*innen Fragebogen (Selbsteinschätzung)
Kursleiter*innen Beobachtungsbogen (Fremdeinschätzung)	Kursleiter*innen Beobachtungsbogen (Fremdeinschätzung)

Eine Übersicht aller erhobenen Variablen ist im Anhang zu finden (s. Anhang_1).

3.1 Stichprobe

Die Stichprobe wurde aus den Schülern*innen der Ichoschule der Jahrgangsstufen 5 – 8 gebildet. Alle Teilnehmenden sind Schüler*innen im gebundenen Ganztags, die im Schuljahr 2020/21 an „respect U“-Kursen teilgenommen haben. Einige Schüler*innen haben bereits in vergangenen Schuljahren am Programm teilgenommen.

Eingrenzung der Stichprobe

Für die finalen Berechnungen zu Messzeitpunkt 2 wurde die Auswahl auf diejenigen Schüler*innen reduziert, für welche zu beiden Messzeitpunkten, neben den programminternen Daten mindestens die Selbst- oder die Fremdbeurteilung vorlag. Ausgeschlossen wurden folglich Jugendliche, bei welchen nur zu einem Erhebungszeitpunkt die Selbst- und/oder die Fremdeinschätzung vorhanden war. Als Gründe dafür können potenzielle Klassen- und Wohnortswechsel angeführt werden.

Bereinigte Stichprobe

Die abschließend gebildete Stichprobe kann wie folgt beschrieben werden: Die Stichprobe umfasst 85 Schüler*innen, die zu Messzeitpunkt 1 zwischen 10 und 16 Jahre alt waren ($M = 12.11$, $SD = 1.65$). 51% der Teilnehmer*innen sind männlich, 49% weiblich. 39% sind nicht in Deutschland geboren und 87% sprechen zu Hause überwiegend nicht Deutsch. Durchschnittlich besuchen die Schüler*innen zwei Mal pro Woche einen „respect U“-Kurs, davon 34% im Klassenverband und 66% in klassenübergreifenden Gruppen. Für 38% ist es das erste, für 13% hingegen schon das vierte Jahr im Programm.

3.2 Erhebungsinstrumente

Sowohl der Schüler*innen Fragebogen als auch der Kursleiter*innen Beobachtungsbogen wurden von der Studentin bzw. Mitarbeiterin der LMU erstellt. Die Erhebungsinstrumente sind im Anhang des Evaluationsberichts zu finden (s. Anhang_2 & Anhang_3). Ziel dieser war es, langfristig eine Entwicklungsdokumentation und kurzfristig eine Bestandsaufnahme der personalen und sozialen Kompetenzen zu ermöglichen. Bei der Konzeption der Messinstrumente wurde sich an den Empfehlungen von Sedlmeier und Renkewitz (2013, S. 91f.) für das „Erstellen von standardisierten Fragebögen und Tests“ orientiert. Der Erstellungsprozess der Fragebögen ist in Evaluationsbericht 1 (November 2020) ausführlich dokumentiert.

Der Schüler*innen Fragebogen wurde in der Piloterhebung getestet, erneut sprachlich verbessert und kam dann im Oktober 2020 und Juli 2021, gemeinsam mit dem Kursleiter*innen Beobachtungsbogen, zum Einsatz.

Im Anschluss an Messzeitpunkt 1 wurde der Empfehlung von Sedlmeier und Renkewitz (ebd.) nachgegangen, das Verfahren der Faktorenanalyse anzuwenden. Dieses testet, ob die verwendeten Items (= einzelne Fragen) tatsächlich ähnliche Inhalte erfassen und zu einem Faktor (= Kompetenzdimension) zusammengefasst werden können. Die Ergebnisse der Faktorenanalysen bestätigten für beide Fragebögen die vermutete 6-Faktorenstruktur bzw. die sechs zugrundeliegenden, vorab theoretisch erarbeiteten Kompetenzdimensionen (*Selbstwirksamkeit, Selbstwahrnehmung, Selbstkontrolle, Kooperationsfähigkeit, Empathiefähigkeit, Konfliktfähigkeit*). Darüber hinaus wurde, ebenfalls der Empfehlung der Autoren folgend, eine Trennschärfeanalyse durchgeführt. Unter der Trennschärfe eines Items versteht man den Zusammenhang dessen mit dem Gesamtergebnis des Tests (= Kompetenzdimension) (ebd.). Die Trennschärfe soll eine Einschätzung ermöglichen, wie gut eine einzelne Frage zwischen Personen mit niedriger und hoher Merkmalsausprägung trennt. Beispielsweise zwischen Personen mit hoher und mit niedriger *Kooperationsfähigkeit*. Ein trennscharfes Item erlaubt eine Vorhersage darüber, dass Personen mit einer hohen (niedrigen) Merkmalsausprägung bei einem gegebenen Item einen hohen (niedrigen) Wert erzielen. Für alle verwendeten Items ergab sich ein Trennschärfekoeffizient von $r \geq .70$, was als gut einzustufen ist. Abschließend wurde die Reliabilität der verwendeten Skalen bzw. der Messung der einzelnen Kompetenzdimensionen betrachtet. Die Reliabilität ist ein Maß für die Genauigkeit bzw. Verlässlichkeit einer wissenschaftlichen Messung und kann mit Hilfe verschiedener Methoden getestet werden. Im Falle der vorliegenden Evaluation wurde sich für die Variante der „internen Konsistenz“ entschieden. Die interne Konsistenz ist ein Maß dafür, wie die Items einer Skala miteinander zusammenhängen. Je höher der Zusammenhang zwischen den Items, desto „konsistenter“ ist die Skala. Die interne Konsistenz wird dabei mit dem Kennwert Cronbachs Alpha abgebildet (ebd.). Für alle verwendeten Skalen ergab sich ein Cronbachs Alpha von $>.80$, was als gut bis sehr gut ($>.90$) eingestuft werden kann.

Demnach lassen sich die verwendeten Items je zu den sechs theoretisch vermuteten Kompetenzdimensionen zusammenfassen. Darüber hinaus weisen sie eine als gut zu bewertende Trennschärfe auf und die aus ihnen gebildeten Skalen sind durch ein gutes bis sehr gutes Maß an interner Konsistenz gekennzeichnet.

Diese Kennwerte können dahingehend interpretiert werden, dass die konzipierten Erhebungsinstrumente zielgruppenadäquat sind (1).

3.3 Statistische Auswertung

Die statistische Auswertung beider Messzeitpunkte wurde mit dem Statistikprogramm SPSS durchgeführt. Der vollständige Auswertungsprozess lässt sich in sieben Schritte untergliedern, welche verschiedene Erkenntnisinteressen verfolgen. Diese sollen nun aufgeführt und so ein Bezug zwischen Auswertung und Evaluationsvorhaben geschaffen werden. Im nachfolgenden Ergebnisteil werden allerdings primär die Ergebnisse des zweiten Messzeitpunkts beleuchtet und die Erkenntnisse von Messzeitpunkt 1 (Evaluationsbericht November 2020) nur zusammengefasst dargestellt.

Messzeitpunkt 1 - Schritt 1: Faktorenanalysen

Zu Beginn wurde der bereits erwähnten Empfehlung von Sedlmeier und Renkewitz nachgegangen, das Verfahren der Faktorenanalyse anzuwenden (ebd.). Es wurde überprüft, ob die verwendeten Items tatsächlich ähnliche Inhalte erfassen bzw. zu den theoretisch angenommenen Faktoren zusammengefasst werden können. Auf diese Art und Weise wurde die Struktur des zu konzipierenden Erhebungsinstruments überprüft (1).

Messzeitpunkt 1 - Schritt 2: Deskriptive Statistiken

Anschließend wurden die deskriptiven Statistiken der Selbst- und der Fremdbeurteilungen für die gebildeten Kompetenzbereiche jahrgangsstufenspezifisch analysiert. Dieser Schritt verfolgte das Ziel, dem Evaluationsanspruch der Status- Quo-Analyse gerecht zu werden (2).

Messzeitpunkt 1 - Schritt 3: Korrelationen

Daraufhin wurde untersucht, ob ein positiver Zusammenhang zwischen der jeweiligen Selbst- und Fremdbeurteilung besteht. Ziel dieses Verfahrens war es zu überprüfen, ob ein*e Schüler*in, der*die sich selbst eher positiv/negativ bewertet hat auch von dem*der Kursleiter*in entsprechend bewertet wurde. Dieser Auswertungsschritt verfolgt mehrere Ziele. Bei der Evaluation von Kompetenzen wird empfohlen, das Verfahren der Selbst- mit dem der Fremdbeurteilung zu kombinieren (Frey, 2006). Anhand der Analyse des Zusammenhangs der jeweiligen Beurteilungen können diese in Perspektive gesetzt und die Bestandsaufnahme reflektiert werden. Außerdem kann die These aufgestellt werden, dass ein positiver Zusammenhang der beiden Kompetenzeinschätzungen für die Zielgruppeneignung beider zu konzipierenden Erhebungsinstrumente spricht (1). Abschließend sollte überprüft werden, ob die Selbst- und Fremdbeurteilungen der jeweiligen Kompetenzdimension zu einem übergreifenden Wert zusammengefasst werden können, der in den nachfolgenden Analysen verwendet werden kann.

Messzeitpunkt 1 - Schritt 4: Multiple Regressionen

Mit Hilfe einer multiplen Regression kann untersucht werden, inwiefern eine beobachtete abhängige Variable (Kompetenz) durch mehrere, unabhängige Variablen erklärt werden kann (Field, 2018). Im Kontext der vorliegenden Evaluation wird mit Hilfe dieses Verfahrens analysiert, ob sich die aus der Fachliteratur herausgearbeiteten Einflussfaktoren *Geschlecht, sozialer Status, Migrationshintergrund* und *Dauer* (der bisherigen Programmteilnahme) - aber auch Hintergrundinformationen wie *Alter* und *Jahrgangsstufe* - auf die Kompetenzausprägungen der Teilnehmenden auswirken. Auf diese Art und Weise sollten potenzielle Einflussfaktoren auf den Kompetenzerwerb der Schüler*innen identifiziert und somit spezifische Förderbedarfe ermittelt werden (3).

Messzeitpunkt 2 – Schritt 1: Korrelationen

Anhand der Daten des zweiten Messzeitpunkts wurde ebenfalls überprüft, ob ein positiver Zusammenhang zwischen der jeweiligen Fremd- und Selbsteinschätzung besteht. Primär sollte hier erneut getestet werden, ob die Bewertungen der jeweiligen Kompetenzdimension zu einem übergreifenden Wert zusammengefasst werden können, der für die weiteren Auswertungsschritte verwendet werden kann.

Messzeitpunkt 2 – Schritt 2: T-Test für abhängige Stichproben

Das Verfahren des t-Tests für abhängige Stichproben überprüft, ob die Mittelwerte zweier abhängiger Stichproben verschieden sind. Von "abhängigen Stichproben" respektive "verbundenen Stichproben" wird dann gesprochen, wenn ein Messwert in einer Stichprobe und ein bestimmter Messwert in einer anderen Stichprobe sich gegenseitig beeinflussen. Dies kann in unterschiedlichen Szenarien der Fall sein (UHZ, 2020). In der vorliegenden Evaluation handelt es sich um das Szenario der „Messwiederholung“. Dies bedeutet, dass die Messwerte (zu Messzeitpunkt 1 und 2) von der gleichen Person stammen, einmal vor und einmal nach einem Schuljahr im Programm „*respect U*“. Mit t-Tests für abhängige Stichproben kann demnach überprüft werden, ob sich die Mittelwerte einer Stichprobe vor und nach einer Intervention signifikant voneinander unterscheiden (ebd.). Mit Hilfe dieses Auswertungsschritts sollte das Erreichen der Projektziele überprüft werden. Es wurde auf Basis der übergeordneten Evaluationsfragestellungen (s. 2.2) analysiert, ob die Schüler*innen nach einem Schuljahr eine positive Tendenz in den Bereichen der personalen und sozialen Kompetenzen bzw. den sechs Kompetenzdimensionen aufweisen. Diese positive Tendenz zeigt sich dabei in Form einer signifikanten Mittelwerts-Differenz.

Messzeitpunkt 2 – Schritt 3: Multiple Regressionen

Auf Basis der Implikationen der Masterarbeit sollten abschließend potenzielle programmspezifische Einflussfaktoren auf die Kompetenzbewertungen der Teilnehmer*innen am Ende des Schuljahres identifiziert werden. Hier kam erneut das Verfahren der multiplen Regressionsanalyse zum Einsatz. Beispielsweise konnte in bisherigen Analysen anhand der StEG-Daten aufgezeigt werden, dass die Regelmäßigkeit der Programmteilnahme die Wirkung von außerunterrichtlichen Angeboten an Ganztagschulen bedingt (DIPF, 2019). Da alle Schüler*innen regelmäßig am Programm teilnehmen wurde an dieser Stelle, in Absprache mit Gesellschaft macht Schule, hinsichtlich der *Intensität* der Angebotsnutzung differenziert. Von Seiten des Auftraggebers wurde darüber hinaus die These aufgestellt, dass die Zusammensetzung der jeweiligen Modulgruppen bzw. deren Dynamik ebenfalls einen Einfluss auf den Kompetenzerwerb haben könnte. Aus diesem Grund wurde die Variable *Gruppenzusammensetzung* (im Klassenverband / nicht im Klassenverband) in die Analyse mitaufgenommen. Abschließend zeigte eine weitere Studie im Ganztagsbereich auf, dass neben derartigen strukturellen Charakteristika die wahrgenommene *Attraktivität* des Angebots, einen Effekt auf dessen Wirkung haben kann (Linberg, Struck & Bäumer, 2018). D.h. Angebote entfalten eher ihre Wirkung, wenn sie von den Schüler*innen als attraktiv wahrgenommen werden. Aus diesem Grund wurde die Variable *Attraktivität* ebenfalls in die Analyse mit aufgenommen.

4 Ergebnisdarstellung und -interpretation

Im Folgenden werden die Ergebnisse des externen Evaluationsvorhabens präsentiert und interpretiert, wobei der Schwerpunkt der Betrachtung auf dem zweiten Messzeitpunkt liegt.

4.1 Messzeitpunkt 1

Zu Beginn erfolgt eine knappe Ergebnisrekapitulation des ersten Messzeitpunkts. Die nachfolgenden Ergebnisse wurden dem „*respect U*“ Team im Dezember 2021, im Rahmen eines Workshops präsentiert. Es wurde über deren mögliche Umsetzung, im Sinne der prozessbegleitenden Evaluation, diskutiert. Außerdem waren die Ergebnisse des ersten Messzeitpunkts Gegenstand der zu verfassenden Masterarbeit, welche dem Projektteam zusätzlich zur Verfügung gestellt wurde. Auf diese Art und Weise sollte die Entwicklungsfunktion (2.2) der Evaluation erfüllt werden.

4.1.1 Deskriptive Statistiken

Im ersten Schritt wurden die deskriptiven Statistiken der Selbst- und Fremdbeurteilungen der sechs Kompetenzdimensionen jahrgangsstufenspezifisch analysiert. Zu diesem Zwecke wurde ein Soll-Niveau für die jeweiligen Kompetenzeinschätzungen, auf Basis der verwendeten Skala (1 = *stimmt gar nicht*, 2 = *stimmt wenig*, 3 = *stimmt manchmal*, 4 = *stimmt total*) bestimmt - im Rahmen derer ein durchschnittlicher Wert von bzw. ab 2.5 für eine positive Einschätzung der Kompetenzausprägung spricht. Anhand der Differenzierung nach Jahrgangsstufen lassen sich deskriptive Unterschiede, sowohl zwischen den einzelnen Kompetenzbereichen als auch zwischen den verschiedenen Jahrgangsstufen, nachvollziehen. Die Statistiken der einzelnen Jahrgangsstufen sind im Anhang des Berichts zu finden (s. Anhang_4). In Hinblick auf die einzelnen Jahrgangsstufen kann festgehalten werden, dass die Schüler*innen aus höheren Klassenstufen zu Beginn des Schuljahres überwiegend höhere Kompetenzwerte aufwiesen.

In Bezug auf die verschiedenen Kompetenzbereiche wird deutlich, dass in der Kategorie der personalen Kompetenzen primär die Dimension *Selbstkontrolle* kritisch bewertet wurde. Im Bereich der sozialen Kompetenzen wurde die Dimension *Konfliktfähigkeit* eher kritisch beurteilt, wohingegen beispielsweise die Dimension der *Kooperationsfähigkeit* durchgehend sehr positiv eingeschätzt wurde.

Demnach wurde im Sinne der prozessbegleitenden Evaluation die Implikation abgeleitet, dass die Kursleiter*innen anhand der jahrgangsstufenspezifischen Kompetenzeinschätzungen die Förderschwerpunkte ihrer Kursgestaltung spezifischer setzen sollten. Beispielsweise könnten vermehrt Programmelemente eingesetzt werden, welche auf die Förderung von *Selbstkontrolle* oder *Konfliktfähigkeit* in Form von Rollenspielen abzielen, anstatt den Schwerpunkt auf Teamspiele bzw. *Kooperationsfähigkeit* zu setzen.

4.1.2 Korrelationen

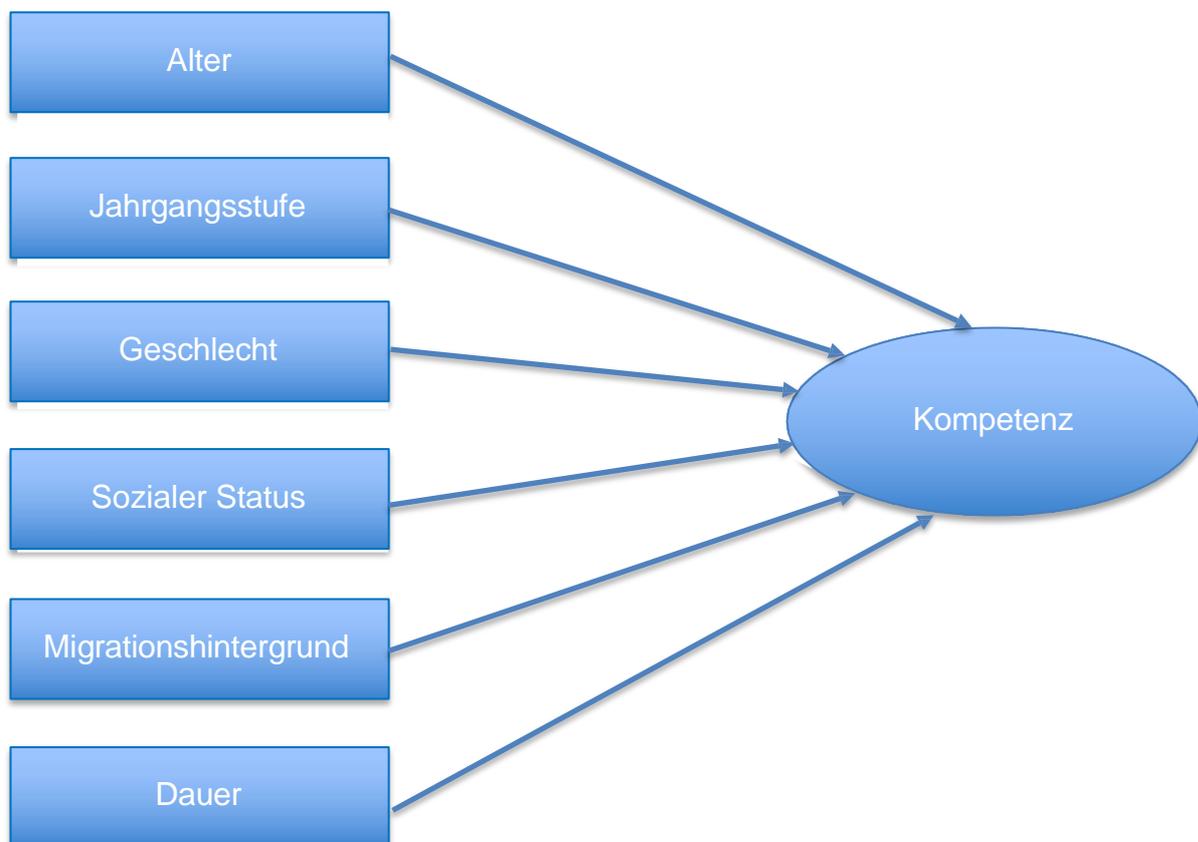
Es konnte gezeigt werden, dass in allen Kompetenzbereichen ein schwacher positiver Zusammenhang zwischen den Angaben der Schüler*innen und der, der Kursleiter*innen besteht. In fünf von sechs Fällen (Ausnahme: *Empathiefähigkeit*) ist dieser Zusammenhang sogar signifikant. Das Signifikanzniveau wurde hier auf $p < .05$ festgesetzt. Auf Basis der Ergebnisse der Trennschärfe-, Reliabilitäts- und Faktorenanalysen sowie der Korrelationen wurden die konzipierten Erhebungsinstrumente als zielgruppenadäquat bewertet (1).

Im Anschluss an die Berechnung der Korrelationen wurden die signifikant korrelierenden Selbst- und Fremdbeurteilungswerte zu einem übergreifenden Wert pro Kompetenzdimension je Schüler*in zusammengefasst. D.h. für jede*r Schüler*in lag nun eine übergreifende Kompetenzeinschätzung pro -bereich vor. Die Ausnahme bildeten die Beurteilungen der *Empathiefähigkeit*, welche aufgrund des nicht- signifikanten Ergebnisses nach wie vor getrennt voneinander zu betrachten waren.

4.1.3 Multiple Regressionen

Das dritte Ziel der Evaluation war es, Einflussfaktoren bzw. spezifische Förderbedarfe hinsichtlich der Kompetenzausprägungen der Schüler*innen zu identifizieren. Im letzten Auswertungsschritt des ersten Messzeitpunkts wurde deshalb das Verfahren der multiplen Regressionsanalyse angewandt.

Anhand dessen sollte überprüft werden, ob die auf Basis von Fachliteratur herausgearbeiteten Prädiktoren (*Alter, Jahrgangsstufe, Geschlecht, sozialer Status, Migrationshintergrund* und *Dauer*) einen signifikanten Einfluss auf die jeweiligen Kompetenzbeurteilungen haben bzw. für welche Subgruppen der Schüler*innen verstärkte Förderbedarfe identifiziert werden können.



Die kompetenzspezifischen Ergebnisse dieses Analyseschritts sind im Anhang des Evaluationsberichts zu finden (s. Anhang_5).

Übergreifend können die Einflussfaktoren des *Geschlechts*, des *Alters* und der *Jahrgangsstufe*, aber auch des *sozialen Status* benannt werden. Dies bedeutet, dass die Kompetenzausprägungen von älteren Schüler*innen, die eine höhere Jahrgangsstufe besuchen, mehrheitlich weiblich sind sowie über einen höheren sozialen Status verfügen (gemessen am Berufsprestige der Eltern), besser bewertet wurden. Im Umkehrschluss bilden die jüngeren männlichen Schüler, welche die unteren Jahrgangsstufen besuchen und über einen geringeren sozialen Status verfügen die zentrale Förderbedarfsgruppe innerhalb der Teilnehmenden.

An dieser Stelle gilt es einen zentralen Unterschied zwischen den Selbst- und den Sozialkompetenzen herauszuarbeiten. Während das *Geschlecht* und der *soziale Status* innerhalb beider Kompetenzkategorien als Einflussfaktoren identifiziert werden konnten, spielt darüber hinaus bei den Selbstkompetenzen überwiegend das *Alter* und bei den Sozialkompetenzen überwiegend die *Jahrgangsstufe* der Teilnehmenden eine Rolle. Dies kann dahingehend interpretiert werden, dass Sozialkompetenzen im Rahmen der Klassen- und Schulgemeinschaft (soziale Umwelt) und Selbstkompetenzen im individuellen Entwicklungsprozess entwickelt werden.

Zudem gilt es auf den Umstand einzugehen, dass für die Dimension *Kooperationsfähigkeit* keine signifikanten Prädiktoren ermittelt werden konnten. Eine Erklärung, weshalb keine signifikanten Einflussfaktoren für diesen Kompetenzbereich identifiziert werden konnten, ist die Varianzeinschränkung der abhängigen Variablen. Die Kompetenzausprägungen der Schüler*innen variieren an dieser Stelle kaum und werden durchgehend sehr gut bewertet. Dies ist an den deskriptiven Befunden erkennbar und erschwert die statistische Identifikation von Einflussfaktoren.

Die deskriptiven Befunde können ebenfalls für die Erklärung des nicht signifikanten Einflusses der Variable *Dauer* auf die Kompetenzwerte herangezogen werden. Hier gilt es anzumerken, dass das Programmjahr 1 zu Messzeitpunkt 1 mit 39% aller Teilnehmenden über- und das Programmjahr 4 mit nur 13% der Teilnehmenden unterrepräsentiert war. Deskriptiv lässt sich ein Einfluss der unabhängigen Variablen vermuten, dieser wird jedoch nur im Falle der Fremdeinschätzung der *Empathiefähigkeit* signifikant. Kritisch zu hinterfragen ist allerdings die, anhand der ungleichen Verteilung der Teilnehmenden auf die Programmjahre zu erkennende, Fluktuation im Projekt.

4.1.4 Zusammenfassung Messzeitpunkt 1

Im Folgenden sollen die Ergebnisse des ersten Messzeitpunkts knapp zusammengefasst werden.

Auf Basis der Ergebnisse der deskriptiven Statistiken kann festgehalten werden, dass die Schüler*innen der höheren Jahrgangsstufen zu Beginn des Schuljahres sozial – und personal kompetenter bewertet wurden als die Teilnehmenden der unteren Klassen. Daraus kann der Förderbedarf der jüngeren Teilnehmer*innen im Programm abgeleitet und die Notwendigkeit der vermehrten wöchentlichen Projektstunden in Jahrgangsstufe 5 (drei- statt zwei) bestätigt werden. Außerdem kann herausgearbeitet werden, dass über alle Jahrgangsstufen hinweg die Kompetenzdimensionen der *Selbstkontrolle* und der *Konfliktfähigkeit* am schlechtesten bewertet wurden. Aus diesem Grund kann hier die Implikation für das Projekt herausgearbeitet werden, diese Kompetenzen im Rahmen von „*respect U*“ verstärkt zu fördern.

Die Ergebnisse der Korrelationen zeigen, dass die Kompetenzeinschätzungen der Schüler*innen mit denen der Kursleitungen in fünf von sechs Fällen (Ausnahme: *Empathiefähigkeit*) signifikant übereinstimmen. D.h., dass die Teilnehmenden sich selbst ähnlich bewerten wie die Kursleiter*innen sie einschätzen. Dieser Befund spricht für die Zielgruppeneignung der jeweiligen Erhebungsinstrumente, da beide befragten Gruppen ein ähnliches Verständnis von den gestellten Fragen haben.

Abschließend können die Befunde der multiplen Regressionen dahingehend gedeutet werden, dass das *Geschlecht*, das *Alter*, die *Jahrgangsstufe* und der *soziale Status* als Einflussfaktoren auf die Kompetenzbewertungen der Schüler*innen benannt werden können. Dieser Einfluss gestaltet sich dahingehend, dass die jüngeren männlichen Schüler, welche die unteren Jahrgangsstufen besuchen und über einen geringeren sozialen Status verfügen (gemessen am Berufsprestige der Eltern) am schlechtesten bewertet wurden. Sie bilden die zentrale Förderbedarfsgruppe innerhalb der Teilnehmenden des Projekts und sollten in den Fokus der Kursleiter*innen rücken. Im Gegenzug werden die älteren Schüler*innen, die eine höhere Jahrgangsstufe besuchen, mehrheitlich weiblich sind sowie über einen höheren sozialen Status verfügen, besser bewertet.

4.2 Messzeitpunkt 2

Zum Abschluss des vierten Kapitels werden nun die Befunde des zweiten Messzeitpunkts präsentiert.

4.2.1 Korrelationen

Zu Messzeitpunkt 2 konnte in allen sechs Kompetenzbereichen ein schwacher positiver Zusammenhang zwischen den Selbst- und Fremdeinschätzungswerten der Kompetenzdimensionen aufgezeigt werden.

Im Unterschied zu Messzeitpunkt 1 war dieser sogar in allen sechs Fällen signifikant. Das Signifikanzniveau wurde dabei erneut auf $p < .05$ festgesetzt.

Daraufhin wurden die signifikant korrelierenden Selbst- und Fremdbeurteilungswerte erneut zu einem übergreifenden Wert pro Dimension je Teilnehmer*in zusammengefasst. D.h. für jede*r Schüler*in lag nun für beide Messzeitpunkte eine übergreifende Kompetenzeinschätzung pro -bereich vor. Ausgenommen blieb trotz des signifikanten Ergebnisses zu Messzeitpunkt 2, die Dimension der *Empathiefähigkeit*. Aufgrund des Umstands, dass die Kompetenzeinschätzungen dieses Bereichs zu Messzeitpunkt 1 getrennt voneinander zu analysieren waren, galt es diese Betrachtungsweise zu Messzeitpunkt 2 simultan fortzuführen, um einen Vergleich (s. 4.2.2) der beiden Zeitpunkte zu ermöglichen.

4.2.2 T-Test für abhängige Stichproben

Im vorletzten Auswertungsschritt des zweiten Messzeitpunkts wurde das Verfahren des t-Tests für abhängige Stichproben angewandt. Ziel dessen war es, zu überprüfen, ob die Teilnehmer*innen nach einem Schuljahr eine positive Tendenz in den Bereichen der Sozial- und Selbstkompetenzen bzw. der sechs bestimmten Subkategorien aufweisen. Dies wurde mittels der Betrachtung der Mittelwertsdifferenzen der beiden Messzeitpunkte analysiert. Auf diese Art und Weise sollten die Evaluationsfragestellungen beantwortet und die Kontrollfunktion der Evaluation erfüllt werden (s. 2.2). Auch hier wurde das Signifikanzniveau auf $p < .05$ festgesetzt.

Selbstwirksamkeit

Hinsichtlich der Einschätzung der *Selbstwirksamkeit* der Teilnehmer*innen kann ein signifikanter Unterschied zwischen den Beurteilungen zu Beginn und zum Ende des Schuljahres aufgezeigt werden ($t = -2.178$, $p = .032$, $n = 85$). Am Ende des Schuljahres wurden die Schüler*innen signifikant besser in Bezug auf ihre *Selbstwirksamkeit* eingeschätzt ($M = 3.13$, $SD = 0.58$) als zu Beginn des Schuljahres ($M = 2.97$, $SD = 0.52$). Die Effektstärke nach Cohen (1992) liegt hier bei $r = .24$ und entspricht damit einem schwachen Effekt.

Selbstwahrnehmung

In Bezug auf die Beurteilung der *Selbstwahrnehmung* der Schüler*innen kann ebenfalls von einem signifikanten Unterschied der Werte vor und nach einem Jahr im Projekt „*respect U*“ gesprochen werden ($t = -2.269, p = .026, n = 85$). Auch hier schnitten die Probanden in den Beurteilungen zu Messzeitpunkt 2 ($M = 3.19, SD = 0.57$) signifikant besser ab als zu Messzeitpunkt 1 ($M = 3.01, SD = 0.63$). Die Effektstärke beträgt $r = .25$, was einem schwachen Effekt entspricht (ebd.).

Selbstkontrolle

Auch im Rahmen der Betrachtung der letzten personalen Kompetenzdimension, der *Selbstkontrolle*, kann eine signifikante Differenz zwischen den beiden Messzeitpunkten identifiziert werden ($t = -4.959, p < .001, n = 85$). Die Bewertungen dieser Kompetenzdimension fielen zu Beginn der „*respect U*“-Kurse signifikant schlechter aus ($M = 2.57, SD = 0.69$) als nach dem Projektjahr ($M = 3.00, SD = 0.66$). An dieser Stelle kann von einem starken Effekt gesprochen werden ($r = .64$) (ebd.).

Kooperationsfähigkeit / Empathiefähigkeit (Selbstwahrnehmung) / Empathiefähigkeit

(Fremdwahrnehmung)

In Bezug auf die *Kooperationsfähigkeit*, die *Empathiefähigkeit* (Selbstwahrnehmung) sowie die *Empathiefähigkeit* (Fremdwahrnehmung) liegen keine signifikanten Mittelwerts-Differenzen zwischen dem Start und dem Ende des Projektjahres vor. Allerdings lässt sich die angestrebte positive Entwicklungstendenz innerhalb aller drei Kompetenzdimensionen deskriptiv erkennen ($t = -1.123, p = .265, n = 85$), ($t = -1.470, p = .146, n = 85$), ($t = -1.283, p = .203, n = 85$)).

Konfliktfähigkeit

Abschließend kann hinsichtlich der Einschätzung der *Konfliktfähigkeit* der Teilnehmer*innen ein signifikanter Unterschied zwischen den Beurteilungen zu Beginn und zum Ende des Schuljahres aufgezeigt werden ($t = -2.869, p = .005, n = 85$). Am Ende des Schuljahres wurde die Konfliktfähigkeit der Schüler*innen signifikant besser bewertet ($M = 3.08, SD = 0.57$) als am Anfang ($M = 2.85, SD = 0.62$). Die Effektstärke nach Cohen (ebd.) beträgt $r = .33$ und entspricht damit einem mittleren Effekt.

Übergreifend

In Rückbezug auf die Evaluationsfragestellungen (s. 2.2) und die herausgearbeiteten Qualitätsindikatoren (s. 2.4) kann demnach festgehalten werden, dass die teilnehmenden Jugendlichen nach einem Schuljahr eine positive Tendenz in dem Bereich der Selbstkompetenzen zeigen. Für alle drei zu betrachtenden personalen Kompetenzdimensionen können signifikant positivere Beurteilungswerte am Ende des Programmjahres verzeichnet werden. Hinsichtlich des Bereichs der sozialen Kompetenzen kann diese Tendenz in den Dimensionen der *Kooperations-* und der *Empathiefähigkeit* deskriptiv erkannt werden – signifikant ist diese allerdings nur im Rahmen der Beurteilung der *Konfliktfähigkeit*. Im Sinne der Kontrollfunktion der externen Evaluation (s. 2.2) kann übergreifend festgehalten werden, dass die für das Projekt aufgewendeten Mittel, ihrem Aufwand gerecht werden.

4.2.3 Multiple Regressionen

Im letzten Auswertungsschritt der Evaluation wurde überprüft, ob die programmspezifischen Einflussfaktoren der *Intensität*, der *Gruppenzusammensetzung* und der wahrgenommenen *Attraktivität* des Angebots, einen Einfluss auf die Kompetenzbeurteilungen nach einem Jahr im Programm haben. Auf diese Art und Weise sollte ein Beitrag zur Lernfunktion der Evaluation (s. 2.2) geleistet werden.

Selbstwirksamkeit / Selbstwahrnehmung / Kooperationsfähigkeit / Empathiefähigkeit (Fremdwahrnehmung) / Konfliktfähigkeit

Die Ergebnisse der Regressionsanalysen zeigten, dass die Intensität des Angebots als signifikanter Einflussfaktor auf die Beurteilung der Kompetenzdimensionen *Selbstwirksamkeit*, *Selbstwahrnehmung*, *Kooperationsfähigkeit*, *Empathiefähigkeit* (Fremdwahrnehmung) und *Konfliktfähigkeit* identifiziert werden kann ($(F(1,78) = 4.38, p = .040, n = 80)$ ($(F(1,78) = 4.88, p = .030, n = 80)$ ($(F(1,78) = 4.79, p = 0.32, n = 80)$ ($(F(1,78) = 8.15, p = .005, n = 80)$ ($(F(1,78) = 8.74, p = .004, n = 80)$). Schüler*innen, die an zwei statt an drei Tagen in der Woche „*respect U*“ besuchen, weisen am Ende des Schuljahres signifikant bessere Kompetenzbeurteilungen in diesen fünf Dimensionen auf. An dieser Stelle ist es wichtig zu erläutern, dass ausschließlich die Schüler*innen der fünften Klassen an drei Tagen in der Woche „*respect U*“ besuchen, weshalb dieses Ergebnis in erster Linie für den zentralen Förderbedarf der unteren Jahrgangsstufen spricht. Diese Interpretation wird im untenstehenden Teilabschnitt „Übergreifend“ detailliert ausgeführt.

Selbstkontrolle / Empathiefähigkeit (Selbstwahrnehmung)

Für die Kompetenzbereiche der *Selbstkontrolle* und der *Empathiefähigkeit* (Selbstwahrnehmung) konnten keine programmspezifischen Einflussfaktoren identifiziert werden.

Übergreifend

Es kann folglich festgehalten werden, dass allein die *Intensität* des Angebots als signifikanter Prädiktor identifiziert werden konnte. Hier gilt es auf die Richtung (negativ) des Einflusses dieser unabhängigen Variablen einzugehen. Die Teilnehmenden, die an zwei statt an drei Tagen in der Woche „*respect U*“ besuchen, wiesen am Schuljahresende höhere Kompetenzwerte auf. Wie bereits erläutert, handelt es sich dabei um die Schüler*innen der höheren Jahrgangsstufen bzw. die älteren Teilnehmenden. Es besteht aufgrund der Organisationsstruktur des Programms ein Zusammenhang zwischen der Anzahl der besuchten Kursstunden pro Woche und der Klassenstufe der Befragten. Aus diesem Umstand heraus kann wiederum die Schlussfolgerung abgeleitet werden, dass der zentrale Prädiktor für die Kompetenzausprägungen, trotz der variierenden Intensität des Angebots, die besuchte *Jahrgangsstufe* bzw. das damit einhergehende *Alter* der Schüler*innen ist. Im Zuge der Implikationen für das Projekt kann hier erneut der zentrale Förderbedarf der unteren Klassenstufen herausgearbeitet werden.

Im Anschluss gilt es den Umstand zu beleuchten, dass die Variablen der *Gruppenzusammensetzung* sowie der *Attraktivität* des Angebots für keine der Kompetenzdimensionen als signifikante Einflussfaktoren identifiziert werden konnten. In Hinblick auf den Einflussfaktor der *Gruppenzusammensetzung* muss an dieser Stelle die Operationalisierung der Variablen kritisiert werden. Die *Gruppenzusammensetzung* (im Klassenverband/nicht im Klassenverband) hängt ebenfalls mit der *Jahrgangsstufe* zusammen und deckt sich im Zuge dessen mit der Variablen *Intensität*. Ausschließlich die Schüler*innen der fünften Klasse, welche dreimal wöchentlich am Programm teilnehmen, tun dies im Klassenverband. Aus diesem Grund kann hier kein neuer Erklärungsbeitrag geleistet werden und die Variable ist in dieser Form redundant. Für zukünftige Evaluationsbestrebungen wird deshalb angeregt, die Variable anders zu operationalisieren, beispielsweise in Form von Fragen zur *Gruppendynamik* statt zur *Gruppenzusammensetzung*. Die Kritik an der Operationalisierung gilt es bei der Betrachtung des Prädiktors der *Attraktivität* des Angebots fortzuführen. Diese wurde zu Beginn des Schuljahres mit Hilfe einer Frage gemessen. Zu der Optimierung der Evaluation im Zuge weiterer Erhebungen wird deshalb angeregt, an dieser Stelle auf bestehende validierte Skalen zum Thema Angebotsattraktivität zurückzugreifen. Dies würde zu einer genaueren Messung des Konstrukts beitragen und die Variable müsste nicht, wie im vorliegenden Fall, für den Einbezug in die Analysen umcodiert werden (*attraktiv* = ja/nein).

Diese Ergebnisse tragen folglich nur bedingt zur Lernfunktion (s.2.2) der externen Evaluation bei. Auf dieses Fazit sowie auf Implikationen für zukünftige Evaluationsvorhaben wird deshalb im nachfolgenden Evaluationsbericht (Juni 2022) vertieft eingegangen.

4.2.4 Zusammenfassung Messzeitpunkt 2

Im Folgenden sollen die Ergebnisse des zweiten Messzeitpunkts knapp zusammengefasst werden. Die Befunde der Korrelationen können analog zu Messzeitpunkt 1 dahingehend interpretiert werden, dass die Teilnehmenden sich selbst ähnlich in Bezug auf ihre personalen und sozialen Kompetenzen einschätzen - wie die Kursleitungen. Die zuvor erläuterte Zielgruppeneignung beider Erhebungsinstrumente kann folglich nicht nur zu Messzeitpunkt 1, sondern auch zu Messzeitpunkt 2 bestätigt werden.

Anhand der Ergebnisse des t-Tests für abhängige Stichproben können die Evaluationsfragestellungen (s. 2.2) beantwortet werden. Die Befunde zeigen, dass die teilnehmenden Jugendlichen nach einem Schuljahr eine positive Tendenz in dem Bereich der Selbstkompetenzen zeigen. Die Schüler*innen wurden in Bezug auf alle drei personalen Kompetenzdimensionen am Ende des Schuljahres signifikant besser bewertet. Im Hinblick auf die Sozialkompetenzen kann ebenfalls von einer positiven Tendenz in allen drei Kompetenzbereichen gesprochen werden, signifikant wurde diese allerdings nur im Fall der *Konfliktfähigkeit*. In Bezug auf die Kontrollfunktion der externen Evaluation (s. 2.2) kann übergreifend festgestellt werden, dass die für das Projekt aufgewendeten Mittel, ihrem Aufwand gerecht werden.

Abschließend kann aus den Befunden der multiplen Regressionen die Implikation für das Programm abgeleitet werden, die Teilnehmer*innen der unteren Klassen noch verstärkter in den Blick zu nehmen. Trotz des Umstands, dass die Schüler*innen der Jahrgangstufe 5 am häufigsten „*respect U*“ Kurse besuchen, werden sie auch am Ende des Schuljahres noch am schlechtesten bewertet.

5 Limitationen

Im Rahmen des letzten Kapitels werden die Limitationen der präsentierten Evaluation vorgestellt.

5.1 Stichprobe

Das übergeordnete Erkenntnisinteresse der vorliegenden Evaluation war es, die Kompetenzentwicklung der Schüler*innen kausal mit der Förderung durch das Projekt in Verbindung zu bringen. Für dieses Vorhaben wäre es von zentraler Bedeutung gewesen, die zu analysierenden Konstrukte zusätzlich in einer Kontrollgruppe zu erfassen.

Anhand der gebildeten Stichprobe ist es nicht möglich, zu differenzieren, ob die präsentierten Veränderungen in den Kompetenzausprägungen sich trotz oder wegen der Teilnahme am Projekt zeigen. In der Projektschule gäbe es passende Vergleichsklassen, allerdings gilt es hier anzumerken, dass die Einführung einer Kontrollgruppe mit einem zu hohen Arbeitsaufwand für ein derartiges Evaluationsvorhaben verbunden gewesen wäre. Diese Entscheidung wurde bewusst getroffen und mit der Stiftung Gesellschaft macht Schule vorab kommuniziert.

5.2 Methodik

Das Evaluationsvorhaben entwickelte sich im Laufe des Projekts und war nicht in dessen Planungsphase integriert. Deshalb ist die Operationalisierung der Qualitätsindikatoren als eingeschränkt anzusehen.

Die konzipierten Erhebungsinstrumente konzentrieren sich auf die, mit dem Projektteam abgestimmten, zentralen Förderdimensionen des Projekts. Allerdings können die erhobenen Variablen nicht alle durch das Projekt geförderten Bereiche abdecken und sind somit als eine Auswahl anzusehen.

Außerdem wurden die Erhebungsinstrumente eigens für die vorliegende Evaluation konzipiert. Aus diesem Grund ist keine Gegenüberstellung der Kompetenzausprägungen der Teilnehmenden mit einer normierten Vergleichsgruppe möglich, was die Ergebnisinterpretation erschwert. Dennoch wurde sich bewusst für die Erarbeitung eigener Messinstrumente entschieden, da bisherige Evaluationsbestrebungen des Projekts an der mangelnden Zielgruppeneignung bereits vorhandener Fragebögen scheiterten.

5.3 Auswertung

Hinsichtlich der Auswertung der Angaben ist das Kriterium der Auswertungsobjektivität als verletzt anzusehen, sowohl bei der Fremd- als auch bei der Selbstbeurteilung. *„Ein Test erfüllt die Forderung nach Auswertungsobjektivität dann, wenn verschiedene Anwender aufgrund der Antworten eines Probanden zu demselben Testergebnis gelangen“* (Sedlmeier & Renkewitz, 2013, S. 70).

In Bezug auf die Fremdbeurteilung ist die Auswertungsobjektivität dahingehend als verletzt anzusehen, dass jeweils eine Kursleitung pro Schüler*in die Bewertung vornahm. An dieser Stelle muss kritisch hinterfragt werden, inwieweit andere Kursleitungen zu dem gleichen Beobachtungsergebnis gekommen wären. Diese Problematik hätte umgangen werden können, wenn mehrere Beobachtungen pro Schüler*in gemacht – und eine Interrater-Reliabilität berechnet worden wäre. Dies konnte jedoch aufgrund mangelnder zeitlicher und personaler Ressourcen nicht umgesetzt werden.

Um dieser Einschränkung entgegenzuwirken wurde vereinbart, dass die Kursleitungen ihre Beurteilungen informell im Team absprechen.

In Hinblick auf die Selbstbeurteilung ist die Erfüllung des Kriteriums ebenfalls als eingeschränkt anzusehen. Grund hierfür ist die Gestaltung des Schüler*innen Fragebogens (s. Anhang_2). Es wurde sich sowohl bei dem Kursleiter*innen als auch bei dem Schüler*innen Fragebogen für geschlossene Antwortoptionen entschieden.

Allerdings sollten die Schüler*innen ihre ausgewählte Antwortoption nicht ankreuzen, sondern anhand von mehr oder weniger zustimmenden Smileys einzeichnen (beispielhaft in der Instruktion demonstriert). Diese Gestaltungsform wurde bewusst gewählt, da so ein höheres Maß an kognitiver Aktivierung bei den Befragten erwartet wurde. Zudem sollte unerwünschtes Antwortverhalten, in Form von musterartigem Ankreuzen, umgangen werden. An dieser Stelle gilt es ebenfalls zu hinterfragen, ob ein*e andere*r Anwender*in als die Autorin, zu den gleichen Testergebnissen gekommen wäre. Für die Gewährleistung eines möglichst hohen Maßes an Auswertungsobjektivität fertigte die Autorin deshalb eine Auswertungsschablone an, welche bei der Analyse beider Messzeitpunkte zum Einsatz kam. Mit Hilfe der Schablone erfolgte eine systematisierte Codierung der Angaben.

6 Fazit und Ausblick

Im Rahmen der externen Evaluation wurden zielgruppenadäquate Erhebungsinstrumente konzipiert und pilotiert, welche die Ausprägungen der Sozial- und Selbstkompetenzen der Teilnehmenden des Projekts erfassen sollten (1). Mit Hilfe dieser Instrumente wurde eine Status-Quo Analyse zu Beginn des Schuljahres durchgeführt (2) und potenzielle Förderbedarfe der Schüler*innen analysiert (3). Die Ergebnisse des ersten Messzeitpunkts wurden dem Team von „respect U“ im Rahmen eines Workshops präsentiert und über deren mögliche Umsetzung, einer prozessbegleitenden Evaluation entsprechend, diskutiert. Darüber hinaus wurden diese detailliert in der Masterarbeit der Autorin analysiert, welche ebenfalls dem Projektteam zur Verfügung gestellt wurde. Auf diese Art und Weise wurde die Entwicklungsfunktion der Evaluation erfüllt. Außerdem wurden Möglichkeiten der weiteren Evaluation erarbeitet bzw. im Rahmen des zweiten Messzeitpunkts umgesetzt (4).

Es wurde kontrolliert, ob die für das Projekt aufgewendeten Mittel ihrem Aufwand gerecht werden (Kontrollfunktion). Dies erfolgte mittels eines Vorher-Nachher-Vergleichs der Kompetenzbeurteilungen zu Beginn und zum Ende des Schuljahres

(5). Die Ergebnisse dessen zeigen, dass die teilnehmenden Jugendlichen nach einem Schuljahr eine positive Tendenz in den Bereichen der Sozial- und Selbstkompetenzen aufwiesen, welche in vier von sieben Kompetenzdimensionen sogar signifikant war. Somit können die Evaluationsfragestellungen beantwortet (s. 2.2) und festgehalten werden, dass die aufgewendeten Mittel ihrem Aufwand gerecht werden.

Abschließend wurden potenzielle programmspezifische Einflussfaktoren auf den Kompetenzerwerb analysiert, um einen Beitrag zur Lernfunktion zu leisten. An dieser Stelle kann aus den Ergebnissen die Implikation für das Programm abgeleitet werden, die Schüler*innen der unteren Klassen noch verstärkter in den Blick zu nehmen. Die Lernfunktion der Evaluation bildet zudem den zentralen Analysegegenstand des nachfolgenden Evaluationsberichts im Juni 2022. Dieser wird sich auf das sechste und letzte Ziel des Vorhabens fokussieren, die konzipierte Evaluation und deren Ergebnisse vertieft zu reflektieren und Implikationen für das Projekt bzw. dessen weitere Evaluation abzuleiten (6).

7 Quellen

- Bauer, K.O. (2007). Theorie und Methodologie der Evaluation an Schule. In K.-O. Bauer (Ed.), *Pädagogisches Training. Evaluation an Schulen: Theoretischer Rahmen und Beispiele guter Evaluationspraxis* (S.13–48). Weinheim: Juventa.
- Burkard, C., & Eikenbusch, G. (2005). *Praxishandbuch Evaluation in der Schule* (4. Aufl.). Berlin: Cornelsen Scriptor.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- DIPF. (2019). *Individuelle Förderung: Potentiale der Ganztagschule*. Verfügbar unter https://www.dji.de/fileadmin/user_upload/bibs2019/28579_20191129_StEG.pdf [09.10.20].
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5. Aufl.). Los Angeles: Sage.
- Frey, A. (2006). Strukturierung und Methoden zur Erfassung von Kompetenz. *Bildung und Erziehung*, 59 (2), S. 125- 145.
- Hense, J., & Mandl, H. (2003). *Selbstevaluation - Ein Ansatz zur Qualitätsverbesserung pädagogischer Praxis und seine Umsetzung am Beispiel des Modellversuchprogramms SEMIK*. Verfügbar unter Microsoft Word - DB_162endgültig.doc (uni-bremen.de) (03.08.21).
- 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 (Hrsg.), *Spuren der Schulevaluation: Zur Bedeutung und Wirksamkeit von Evaluationskonzepten im Schulalltag* (S. 85–99). Bad Heilbrunn: Klinkhardt.
- Sedlmeier, P., & Renkewitz, F. (2013). *Forschungsmethoden und Statistik. Ein Lehrbuch für Psychologen und Sozialwissenschaftler* (2., aktualisierte und erweiterte Aufl.). Hallbergmoos: Pearson.
- UHZ. (2020). *T-Test für abhängige Stichproben*. Verfügbar unter UZH - Methodenberatung - t-Test für abhängige Stichproben (07.09.21).

8 Anhang

Anhang_1: Variablenübersicht

Anhang_2: Deskriptive Statistiken

Anhang_3: Multiple Regressionen

Anhang_4: Schüler*innen-Fragebogen

Anhang_5: Kursleiter*innen-Beobachtungsbogen

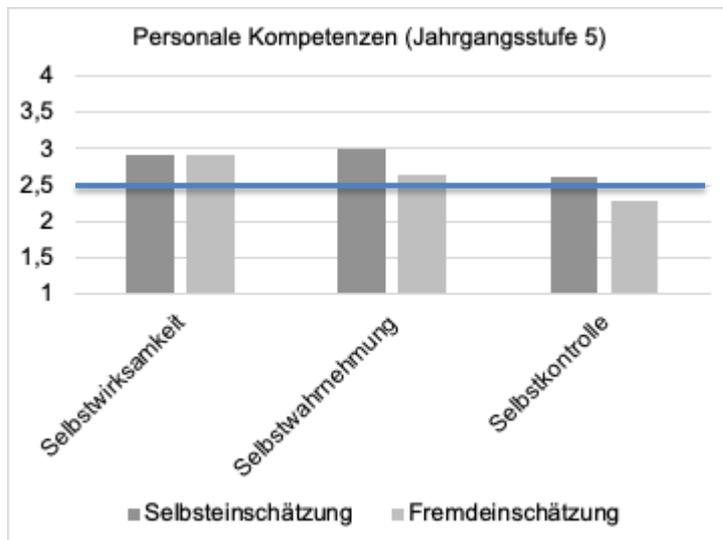
Anhang_1: Variablenübersicht

Dimension	Variable	Quelle	
Hintergrund	Alter	Programminterne Daten	
	Geschlecht		
	Jahrgangsstufe		
	Migrationshintergrund		
	Familiensprache		
	Sozialer Status	Schüler*innen Fragebogen	
Programm- teilnahme	Attraktivität	Programminterne Daten	
	Dauer		
	Intensität		
	Gruppenzusammen- setzung		
Personale Kompetenzen	Selbstwirksamkeit	Schüler*innen Fragebogen	Kursleiter*innen Beobachtungsbogen
	Selbstwahrnehmung		
	Selbstkontrolle		
Soziale Kompetenzen	Kooperationsfähigkeit		
	Empathiefähigkeit		
	Konfliktfähigkeit		

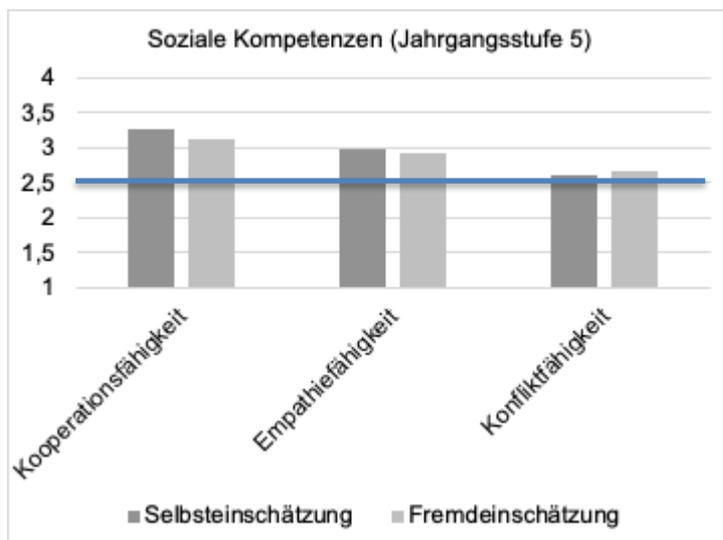
Anhang_2: Deskriptive Statistiken

Jahrgangsstufe 5

Personale Kompetenzen

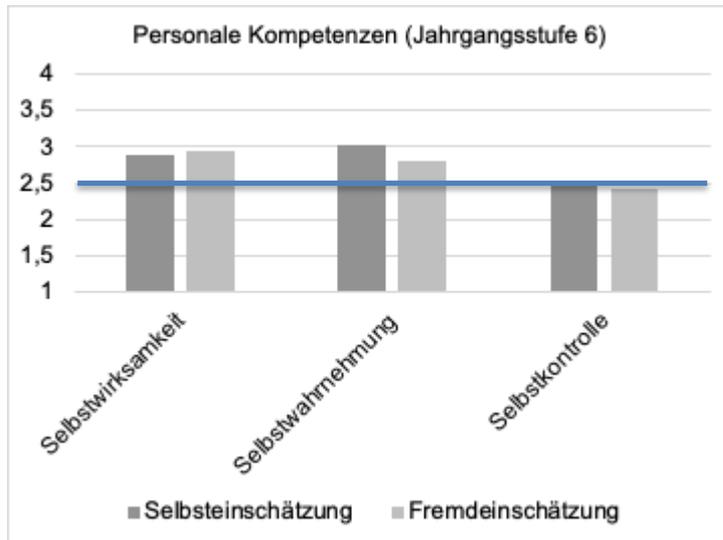


Soziale Kompetenzen

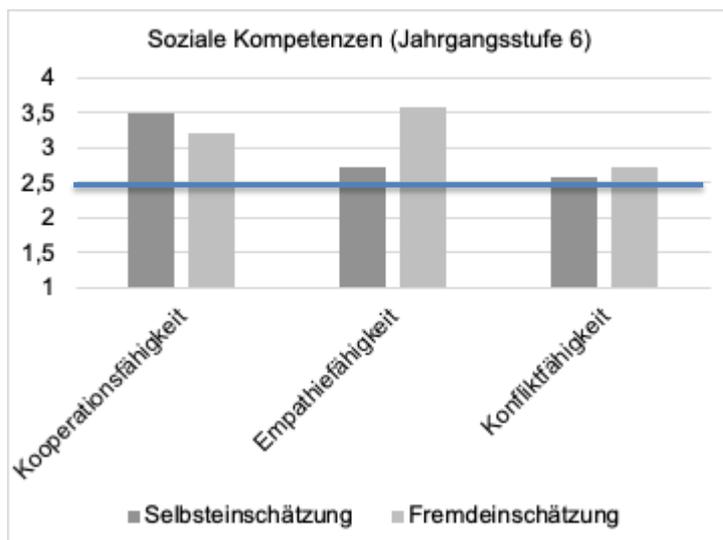


Jahrgangsstufe 6

Personale Kompetenzen

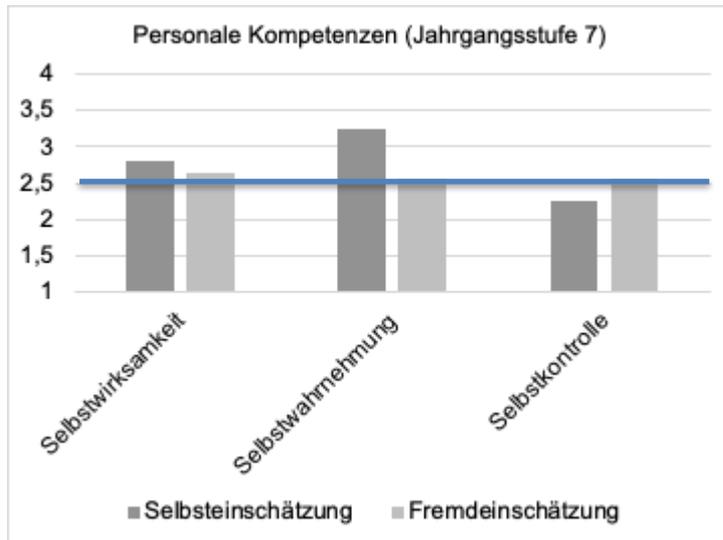


Soziale Kompetenzen

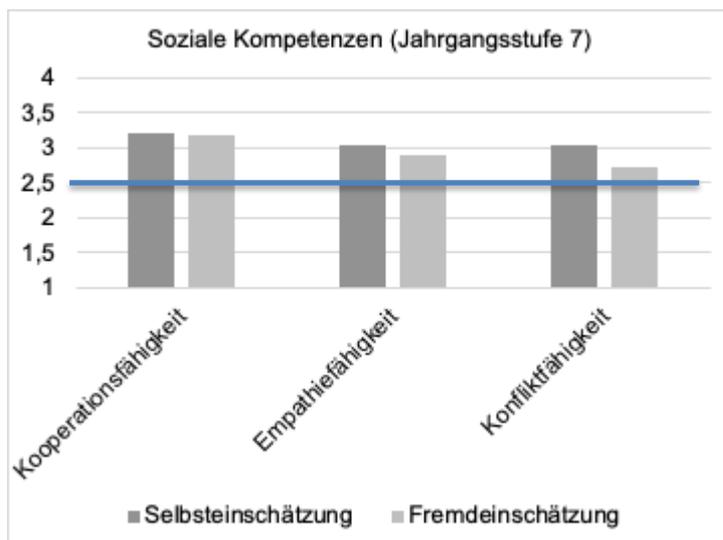


Jahrgangsstufe 7

Personale Kompetenzen

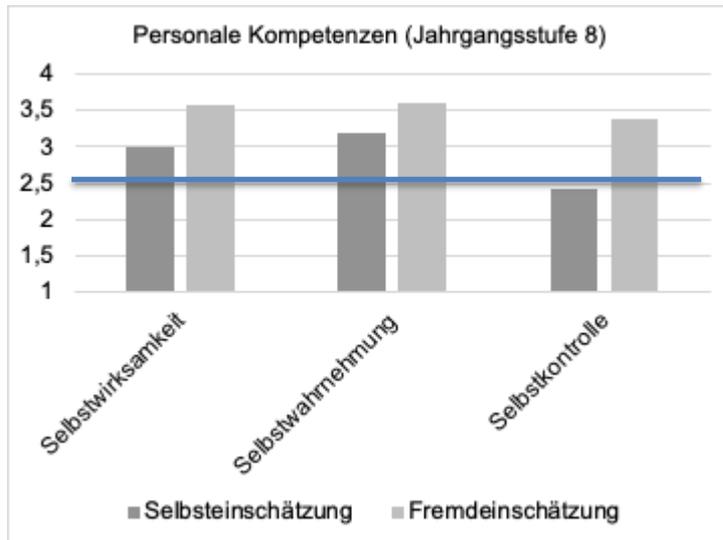


Soziale Kompetenzen



Jahrgangsstufe 8

Personale Kompetenzen



Soziale Kompetenzen

