

Zweite Berichterstattung Schuljahr 24/25

Vergleichsstudie:

Englischkompetenzen und Problemlöseverhalten von Schüler:innen mit CLIL vs. ohne CLIL

Die vorliegende Berichterstattung beinhaltet die Ergebnisse zur Begleitforschung des dritten CLIL-Projektjahres, konkret die Erhebung der Englischkompetenzen und des Problemlöseverhaltens der Schüler:innen durchgeführt am Ende des Schuljahres im Juni 2025.

Zur Erinnerung: Im dritten Projektjahr von 'Buchrain goes International' und des damit verbundenen erweiterten Englischunterrichts im Sinne von CLIL (Content and Language Integrated Learning) wurden die fremdsprachlichen Lernstände der CLIL-Schüler:innen (n=110 Kinder) zugehörig sieben Klassen (3 x 5. Klasse, 4 x 6. Klasse) in allen vier Kompetenzbereichen «Hören», «Lesen», «Sprechen» und «Schreiben» sowie im Bereich des Problemlöse-Verhaltens erhoben (vgl. Kapitel «Hintergrundinformationen» in der ersten Berichterstattung Schuljahr 2024/25). Die durchgeführten Tests decken sich mit den im Lehrplan 21 verlangten Kompetenzniveaus (GER A.1.2.-2.1). Alle Testaufgaben wurden ebenfalls mit fünf Klassen (1 x 5. Klasse, 4 x 6. Klasse,) ohne CLIL durchgeführt (n=88 Kinder). Wiederum wurden die Korrekturen der Tests aus dem Hör- und Leseverstehen mit eindeutigen Lösungen von einer Beurteilerin vorgenommen. Die Bewertungen in den Kompetenzbereichen «Schreiben» und «Sprechen» basieren auf klar definierten Kriterien. Um Konsistenz und Objektivität in den Bewertungen dieser Tests zu garantieren, wurde ein Viertel dieser Bewertungen von zwei Beurteilerinnen vorgenommen.

Die insgesamt 198 beteiligten Kinder wurden in drei Stärkegruppen («high-attainers», «mid-attainers», «low-attainers») eingeteilt. Ausserdem wurden alle Kinder in je eine weitere Untergruppe bezüglich der Kontaktzeit mit Englisch, resp. CLIL zugeordnet: Jene Kinder ohne CLIL («no-CLIL»: 8-10 Wochenlektionen Englisch¹), jene Kinder mit wenig CLIL («low-CLIL»: 10-14 Wochenlektionen Englisch, inkl. 1-2 Lektionen CLIL während 1-2 Schuljahren) und jene Kinder mit mehr CLIL («high-CLIL»: 16-21 Wochenlektionen Englisch, inkl. 2-6 Lektionen CLIL während drei Schuljahren). Mit diesem Wert entsteht eine vergleichbare Grösse – mehr oder weniger unabhängig der Klassenstufe – mit der evaluiert werden kann, inwiefern die unterschiedlichen Kontaktzeiten mit Englisch die Resultate beeinflusste. In der Tabelle 1 sind die Anzahl Kinder zugehörig der Untergruppen, d.h. den Stärkegruppen und Zuordnungen nach Kontaktzeit mit Englisch, resp. CLIL zusammengefasst.

Tabelle 1: Übersicht der Anzahl Kinder aufgeteilt nach Untergruppen

	Low-attainers	Mid-attainers	High-attainers	Total
Non-CLIL (8-10 Wochenlektionen Englisch)	24	32	32	88
Low-CLIL (10-14 Wochenlektionen Englisch/CLIL)	19	26	18	63
High-CLIL (16-21 Wochenlektionen Englisch/CLIL)	13	18	16	47
Total	56	76	66	198

Weil die Vermischung der Klassenstufen (5. und 6. Klasse) unklare Datenstrukturen hervorbrachte, die schwierig zu analysieren und interpretieren waren, wurden auch noch weitere Analysen nur mit den 6. Klasskindern durchgeführt. Dies weil es dort zwei annähernd vergleichbare Gruppen à je vier Klassen gibt: jene Klassen ohne CLIL in Ebikon und jene mit CLIL in Buchrain.

¹ Diese Anzahl berechnet sich aus den Anzahl Englischlernjahren à Wochenlektionen (z.B. 3. Klasse = 3 Lektionen, 4. Klasse = 3 Lektionen, 5. Klasse = 2 Lektionen: Total: 8 Wochenlektionen Englisch)

Tabelle 2: Übersicht der Anzahl 6.Klasskinder aufgeteilt nach Untergruppen

	Low-attainers	Mid-attainers	High-attainers	Total
Non-CLIL (4 x 6. Klassen)	22	23	26	71
CLIL (4 x 6. Klassen)	24	25	15	64
Total	46	48	41	135

Insgesamt adressiert der vorliegende Bericht folgende vier Fragen:

- 1) Welche allgemeinen Einstellungen haben die Lernenden gegenüber dem Englisch und wie schätzen sie sich ein? Welche motivationalen Einstellungen gegenüber CLIL bringen die CLIL-Schüler:innen mit?

Diese Einstellungen wurden durch Selbsteinschätzung mittels Fragebogen erhoben und sollen einen ersten Eindruck geben, was die unterschiedlichen Untergruppen gegenüber der Zielsprache empfinden und wie sie selbst ihre Kompetenzen im Englisch einschätzen. Des Weiteren wurden die CLIL-Kinder befragt, inwiefern sie diese Lektionen gut finden und ob sie sich darauf freuen. Damit lassen sich ansatzweise motivationale Einstellungen erheben.

- 2) Wie schnitten die CLIL-Lernenden im Vergleich zu jenen Kindern ohne CLIL bei den unterschiedlichen Tests ab?

Für die Thematisierung dieser Frage, wurden alle beteiligten Kinder nach CLIL und non-CLIL eingeteilt und die Mittelwerte der Ergebnisse zu jedem Test verglichen. An dieser Stelle muss angemerkt werden, dass die Auswertung der Daten nach dieser Aufteilung etwas problematisch ist. Dies weil die Gruppen mit und jene ohne CLIL sehr heterogen in Bezug auf Anzahl Klassen, folglich Anzahl Kinder, sowie hinsichtlich der Klassenstufen (5. oder 6. Klassen) als auch CLIL-Kontaktzeit zusammengesetzt sind. Ausserdem konnte nicht erhoben werden, inwiefern die Klassen überhaupt insgesamt vergleichbar sind (z.B. in Bezug auf allgemeine Leistungsfähigkeit, Lerndispositionen oder auch sozioökonomische Faktoren). Aus all diesen Gründen, sind deshalb die nachfolgenden weiteren Fragen aussagekräftiger und werden in diesem Bericht mehr Raum gegeben:

- 3) Inwiefern haben sich die Ergebnisse zwischen anfangs und Ende des Schuljahres für jede Untergruppe verändert, d.h. unter Berücksichtigung von Englischleistungen und Kontaktzeit mit Englisch/CLIL?

Weil in diesen Untergruppen auch 5. Klasskinder in unterschiedlichen Anteilen dabei sind und davon auszugehen ist, dass die allgemeine (kognitive) Reife der Schüler:innen und ein Jahr weniger Englisch-(Unterricht) auch einen Einfluss auf die Resultate haben könnte, werden abschliessend auch noch die Resultate der 6. Klasskinder mit und ohne CLIL verglichen.

- 4) Wie haben sich die Ergebnisse bei den vergleichbaren Gruppen an 6. Klasskindern innerhalb eines Schuljahres verändert?

Und ganz zum Schluss, um die Aussagekraft mittels eines grösseren Samplings (d.h. ohne Untergruppen) zu erhöhen und nochmals das wesentliche hervorzuheben, werden die Englischergebnisse insgesamt zusammengefasst. So können auf einen Blick die Fortschritte für das Fach Englisch und bezüglich Problemlösefähigkeiten der CLIL-Schüler:innen vs. jenen ohne CLIL aufgezeigt werden.

- 5) Wie haben sich die Ergebnisse der 6. Klasskindern mit CLIL im Englisch insgesamt und beim «Tower of London» im Vergleich zu jenen ohne CLIL innerhalb eines Schuljahres verändert?

Bei den nachfolgenden Auswertungen werden diese fünf Fragen folgendermassen berücksichtigt: In einem ersten Schritt werden die Einstellungen der Untergruppen deskriptiv dargelegt, dann jeweils die Gruppen CLIL und non-CLIL verglichen, jeweils auch mit Bezug zu den Resultaten des ersten Erhebungszeitpunktes und letztlich werden die differenzierten Ergebnisse für jede Untergruppe dargestellt. Im Anschluss an die Präsentation der Ergebnisse zu jeder getesteten Kompetenz, werden die wichtigsten Erkenntnisse zusammengefasst und bereits vorsichtig interpretiert. Ganz am Schluss werden die gesamten Ergebnisse besprochen und weitere Zusammenhänge sowie Limitationen dieser Untersuchung dargelegt.

Ergebnisse: Einstellungen

Als erstes ist von Interesse zu erfahren, ob sich die Lernenden selbst so einschätzten, wie sie von ihren Lehrpersonen wahrgenommen wurden. Die Schüler:innen wurden mittels Fragebogen gebeten, die Aussagen

«Ich bin gut im Englisch.» auf einer Skala von «Ja.» (4), «Eher ja.» (3), «Eher nein.» (2), «Nein.» (1) und «Ich weiss es nicht.» (0) einzuschätzen. Wie in der Abbildung 1 ersichtlich, stimmen diese Einschätzungen mit den Wahrnehmungen der Lehrpersonen überein, d.h. für jede Intensitätsstufe (non-CLIL, low-CLIL, high-CLIL) sind die Einschätzungen der Schüler:innen mehrheitlich kongruent zu ihren zugeteilten Leistungsstärken. Nur bei der Gruppe mit hohem CLIL-Anteil schätzten sich die starken Schüler:innen etwas weniger positiv ein als ihre Peers. Das könnte darauf hindeuten, dass entweder die lernstarken CLIL-Schüler:innen dazu neigen sich zu tief einzuschätzen im Englisch; oder auch dass die Abgrenzung und Einteilung der Lernenden in die Gruppe der durchschnittlichen und starken Lernenden nicht ganz so deutlich vorgenommen werden konnte. Ebenfalls kann festgestellt werden, dass es einen Zusammenhang zu geben scheint zwischen Selbsteinschätzungen im Englisch und Kontaktzeit mit der Zielsprache. Konkret heisst das, dass je mehr man mit Englisch/CLIL in Kontakt tritt, desto besser schätzen sich die vergleichbaren Leistungsgruppen ein.

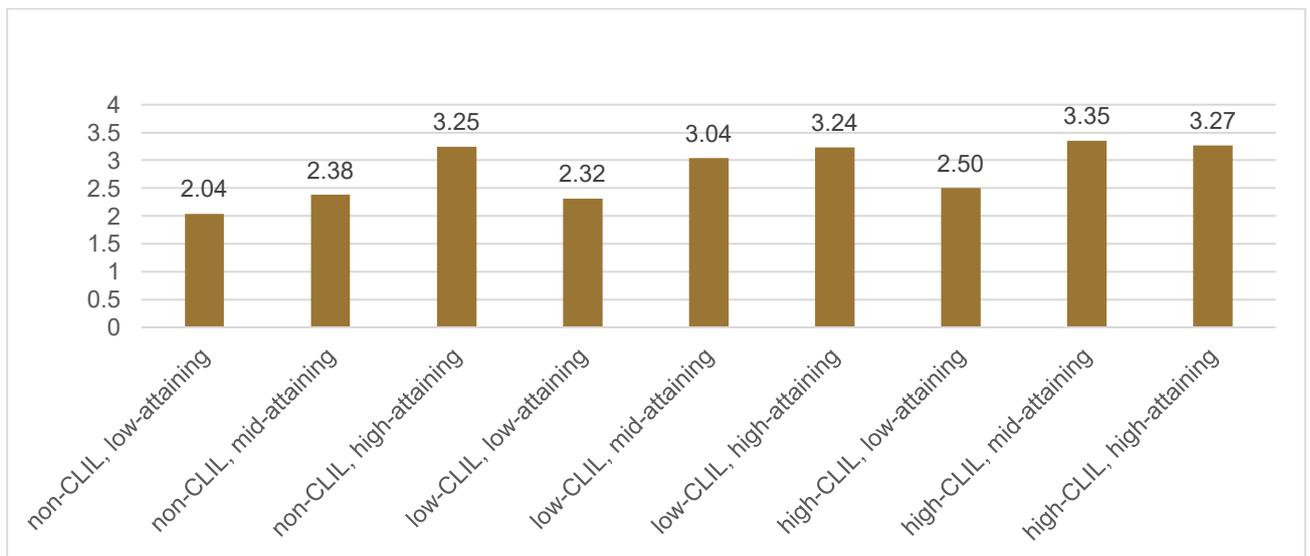


Abbildung 1: Selbsteinschätzungen der Lernenden zum Statement «Ich bin gut im Englisch.» (T2)

Wie in der Abbildung 2 ersichtlich lernen alle Schüler:innen mehrheitlich gerne Englisch. Auch hier ist für jeden Erhebungszeitpunkt zu beobachten – mit einer Ausnahme bei den low CLIL, mid-attainers bei T2 –, dass je höher die die Leistungsgruppe, desto lieber haben sie Englisch. Ebenfalls interessant ist zu sehen, dass die low- und high-attainers mit CLIL deutlich lieber Englisch lernen als jene ohne CLIL. Ebenfalls deutlich zu erkennen, dass die Motivation für das Englisch lernen der Kinder mit CLIL übers Schuljahr – bis auf eine Gruppe – zunahm. Bei den Kindern ohne CLIL nahm die Motivation fürs Englischlernen im Laufe der Zeit ab oder sie blieb unverändert.

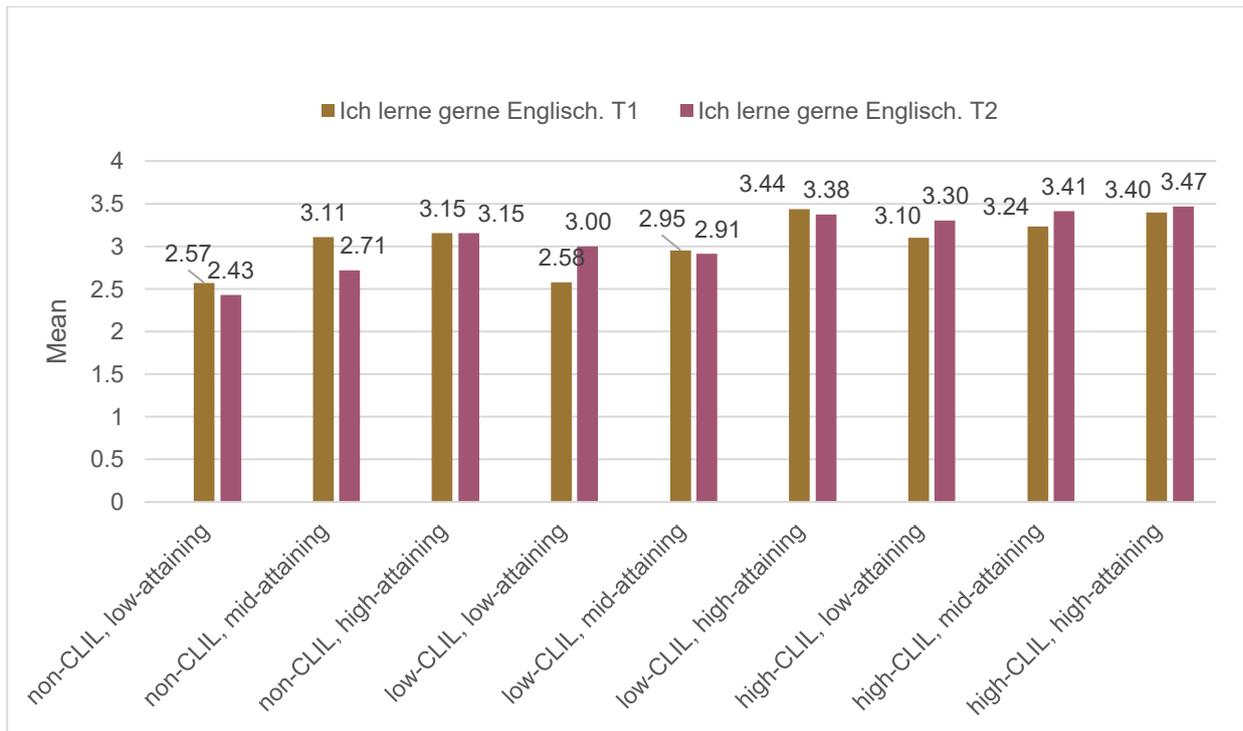


Abbildung 2: Selbsteinschätzungen der Lernenden zum Statement «Ich lerne gerne Englisch.» (T1 vs. T2)

Ob die CLIL-Kinder es gut fanden, dass sie weitere Fächer auf Englisch haben, ist in der Abbildung 3 visualisiert. Auch hier zeigt sich die Tendenz, dass je mehr CLIL und je lernstärker die Lernenden, desto positiver ist ihre Einstellung zu CLIL. Wobei es anzumerken gilt, dass die mid-attainers mit mehr CLIL dieser Unterrichtsform Ende Schuljahr (T2) insgesamt am positivsten gegenüberstanden.

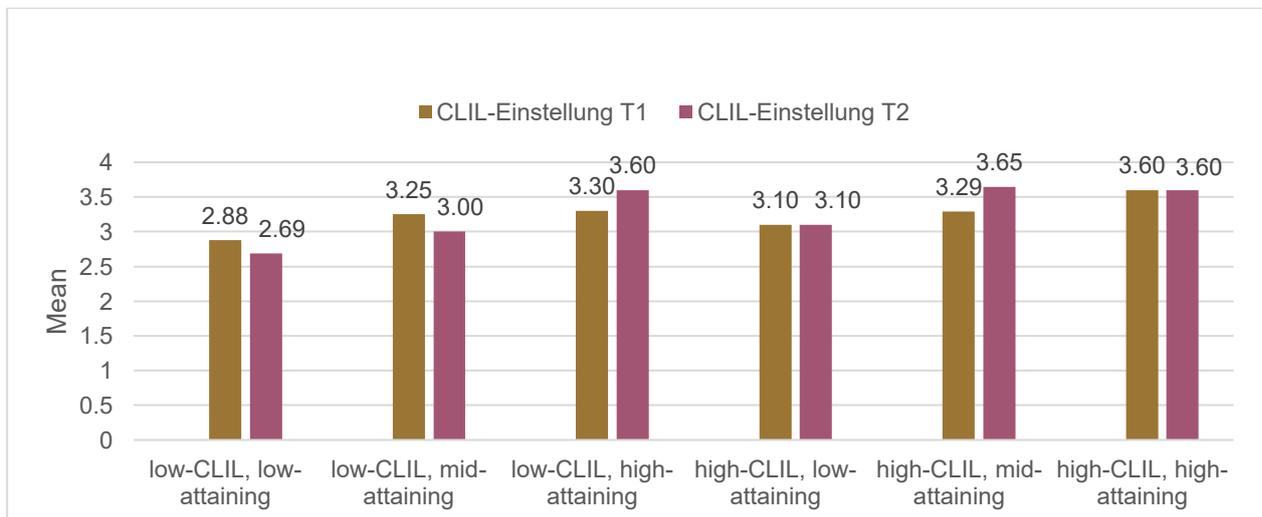


Abbildung 3: Selbsteinschätzungen der CLIL-Lernenden zum Statement «Ich finde es gut, dass wir CLIL haben.» (T1 vs. T2)

Als letztes wurden die Lernenden gefragt, ob sie sich auf die CLIL-Lektionen freuten. Wie in Abbildung 4 ersichtlich, wurden diese persönlichen, wahrscheinlich subjektiv stark gefärbten Einschätzungen weniger kongruent zu den Leistungsstärken vorgenommen. Am meisten freuten sich die Gruppe der mid-attainers mit mehr CLIL, befolgt von beiden Gruppen der high-attainers – die sich am Ende des Schuljahres mehr auf die CLIL-Lektionen zu freuen schienen als noch zu Beginn des Schuljahres. Bei den low-attainers sind es jene mit mehr CLIL, die sich auf diese Lektionen am Ende des Schuljahres mehr freuten als die Peers mit weniger CLIL.

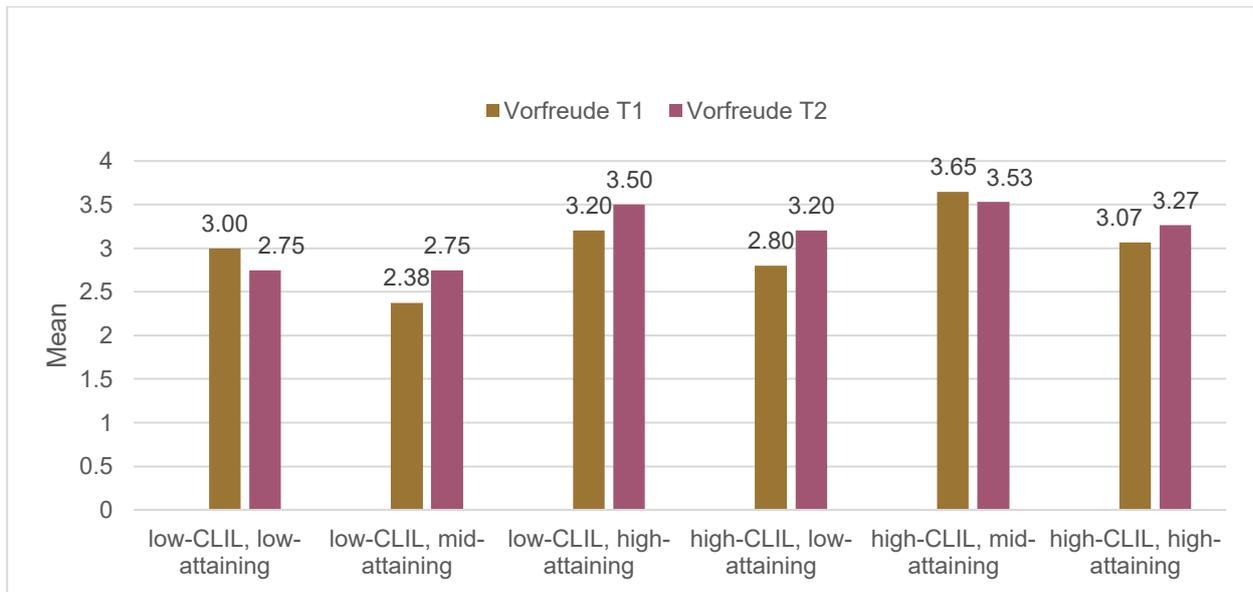


Abbildung 4: Selbsteinschätzungen der CLIL-Lernenden zum Statement «Ich freue mich auf die CLIL-Lektionen.»(T1 vs. T2)

Ergebnisse: Lesen

Im Bereich «Lesen» lösten die Schüler:innen drei alltagsbezogene Aufgaben. Ein Leseverständnis verlangte von den Lernenden das Zuordnen von geplanten Einkäufen in die aufgelisteten Geschäfte in einem «shopping directory» (5 Punkte), eine weitere Aufgabe beinhaltete Buchbesprechungen und deren Bewertungen, ob es sich um positive, neutrale oder negative Aussagen handelte (5 Punkte) und beim letzten Leseverständnis mussten die Lernenden Informationen auf einem Formular kontrollieren (4 Punkte). Maximal konnten 15 Punkte erreicht werden.

In der nachfolgenden Tabelle 3 (links) sind erreichten Durchschnittspunkte der Lernenden mit CLIL jenen ohne CLIL gegenübergestellt, im Vergleich zur ersten Erhebung.

Tabelle 3: Durchschnitt des erreichten Punktetotals im Test «Lesen» der Lernenden ohne und mit CLIL bei T2 vs. T1

T2 (Juni 2025)	Total Punkte (max. 15 P.)	T1 (September 2024)	Total Punkte (max. 15 P.)
non-CLIL	9.64	non-CLIL	9.06
CLIL	9.14	CLIL	8.37

Erfreulich ist zusehen, dass sich die Lernenden bei der Leseaufgabe mit vergleichbaren Schwierigkeitsgrade von der ersten zur zweiten Testdurchführung steigern konnten. Die Lernenden ohne CLIL schnitten wiederum bei der zweiten Erhebung im Leseverstehen mit durchschnittlichen 9.64 Punkten leicht besser ab als jene mit CLIL, die im Durchschnitt 9.14 Punkten erzielten. Der durchgeführte Welch-Test – mit dem überprüft werden kann, ob sich zwei unabhängige Gruppen signifikant unterscheiden und der bei kleinen Stichproben sowie Daten mit ungleicher Varianz robuster ist als ein «normaler» T-Test – lässt keine signifikanten Unterschiede ($t(182) = 1.217, p = 0.227$) zwischen den Lernenden ohne CLIL ($M = 9.64; SD = 2.70$) und jenen mit CLIL ($M = 9.14; SD = 2.76$) erkennen. Der Effekt von 0.18 ist gemäss Cohen (1988)² als klein einzustufen.

Wie diese Daten differenzierter dargestellt nach Untergruppen, d.h. Kontaktzeit mit Englisch/CLIL und Leistungsgruppe aussehen, ist in der Abbildung 5 dargestellt.

² Gemäss Cohen (1988) bedeutet $d < 0.4$ ein schwacher Effekt, $0.5 < d < 0.7$ einen mittleren Effekt, $0.8 < d$ ein starker Effekt.

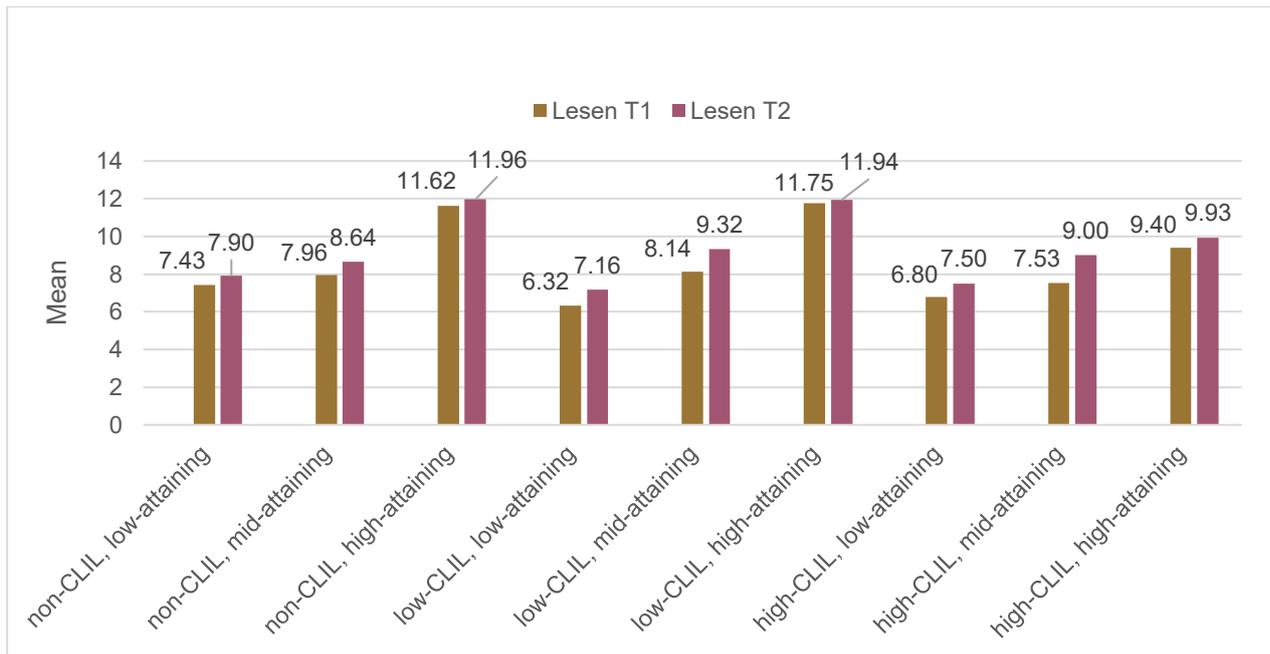


Abbildung 5: Durchschnittliche Ergebnisse im Bereich «Lesen» nach Untergruppen und Erhebungszeitpunkten

Wie in Abbildung 5 ersichtlich konnte auch jede Untergruppe über das Schuljahr hinweg bessere Resultate erzielen. Erneut ist zudem die Punkteverteilung kongruent zu den Leistungsgruppen zu erkennen, d.h. wiederum zeigten die stärkeren Lernenden bessere Resultate beim Leseverständnis als ihre Peers mit tieferen Englischkenntnissen innerhalb der Gruppe mit denselben Kontaktzeiten zu Englisch/CLIL. Die anscheinend starke Kontrollgruppe ohne CLIL zeigte bereits mehrheitlich bessere Resultate beim ersten Erhebungszeitpunkt. Nur die mid-attainers der beiden Gruppen mit CLIL konnten im Vergleich bessere Resultate als die gleiche Stärkegruppe ohne CLIL erzielen. Die grössten Fortschritte scheinen die low- und mid-attainers mit CLIL erzielt zu haben.

Betrachtet man nur die Ergebnisse der 6. Klasskinder, die wie bereits erklärt am ehesten vergleichbar sind, zeigen sich die Resultate im Bereich «Lesen» wie in Abbildung 6 visualisiert. Auffällig ist, dass alle 6.Klasskinder mit CLIL beim zweiten Messzeitpunkt Ende Schuljahr besser abschnitten als die vergleichbaren Stärkegruppen – dies obwohl die low- und mid-attainers ohne CLIL anfänglich stärker waren.

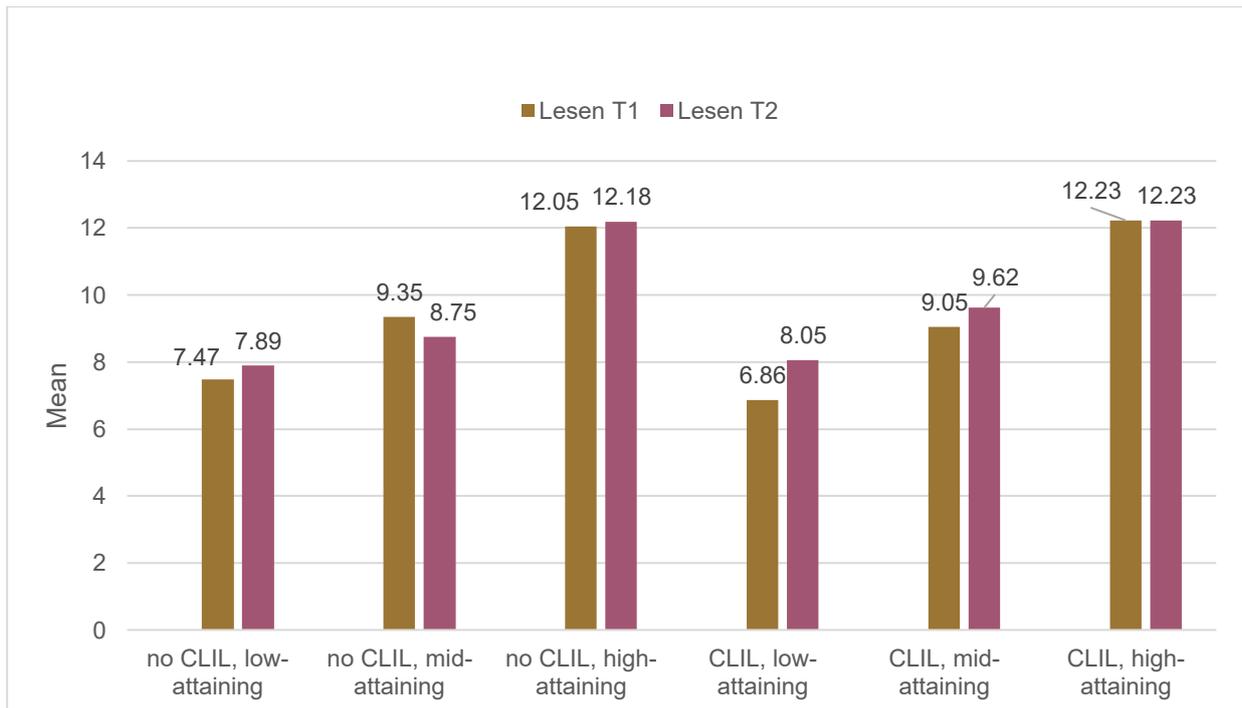


Abbildung 6: Durchschnittliche Ergebnisse der 6. Klassen «Lesen» nach Stärkegruppen und Erhebungszeitpunkten

Da es sich um eine kleine Stichprobe handelt mit nicht normal-verteilten Daten, ist es sinnvoll zusätzlich zum Mittelwert (*mean*), den Median anzuschauen. Der Median, auch Zentralwert genannt, ist der Wert, der genau in der Mitte einer nach Grösse geordneten Datenreihe liegt. Die nachfolgende Tabelle 4 verdeutlicht, dass gemessen am Median z.B. die low- und mid-attainers ohne CLIL am Ende des Schuljahres gleich abschnitten wie am ersten Erhebungszeitpunkt, während sich die gleichen Leistungsgruppen mit CLIL verbesserten. Mit diesem Median in Verbindung stehenden Rangwerten kann mittels Wilcoxon-Test statistisch analysiert werden, ob sich bei zwei abhängigen Stichproben die zentralen Tendenzen unterscheiden (Schwarz et al., 2023). Damit lässt sich in Erfahrung bringen, ob signifikante Ergebnisse zwischen den zwei Messzeitpunkten pro Gruppe vorliegen. Die Resultate zu dieser statistischen Analyse sind in der nachfolgenden Tabelle 4 zusammengefasst.

Tabelle 4: Statistische Werte zum Test «Lesen» der 6. Klasskinder

Untergruppe	Differenz (T1-T2)	Median T1	Median T2	Wilcoxon-Test (exakt)	Effekt (r)
Non-CLIL, low-attainers	0.42	8.00	8.00	$z = -0.599, p = 0.549, n=19$	0.14
Non-CLIL, mid-attainers	-0.60	9.00	9.00	$z = -0.856, p = 0.409, n=20$	0.19
Non-CLIL, high-attainers	0.07	13.00	12.50	$z = -0.692, p = 0.513, n=22$	0.15
CLIL, low-attainers	1.19	7.00	7.50	$z = -2.066, p = 0.038, n=22$	0.44
CLIL, mid-attainers	0.57	9.00	10.00	$z = -0.944, p = 0.345, n=21$	0.21
CLIL, high-attainers	0.00	12.50	12.00	$z = 0.000, p = 1.000, n=13$	0.00

Die statistischen Berechnungen zeigen, dass sich einzig die lernschwachen 6. Klasskinder mit CLIL signifikant steigern konnten. Die Resultate der mid-attainers veränderten sich ebenfalls deutlich, jedoch nicht signifikant: Während sich jene mit CLIL steigern konnten, verschlechterten sich jene ohne CLIL. Die Resultate der high-attainers, die bereits recht hoch ausfielen am ersten Messzeitpunkt, veränderten sich am wenigsten – gemessen am Median verschlechterten sie sich sogar leicht.

Zusammengefasst kann zum Bereich «Lesen» an dieser Stelle festgehalten werden, dass die low-attainers mit CLIL die grössten Fortschritte erzielen konnten. Veranschaulicht mit den Daten der 6. Klasse liegen diese Veränderungen gar im signifikanten Bereich. Dies ist ein erfreuliches Resultat, weil davon auszugehen ist, dass im eher mündlich ausgerichteten CLIL-Unterricht in den handlungsorientierten Fächern das Lesen wahrscheinlich weniger im Fokus steht. Trotzdem scheint der zusätzliche Englischkontakt einen positiven Einfluss auf das Leseverständnis dieser eher lernschwächeren Schüler:innen zu haben. Die high-attainers, die bereits sehr gute Resultate erzielten anfangs Schuljahr konnten sich bei diesen Leseverständnis-Tests mit vorgegebenen Punktemaximum nicht mehr erheblich steigern.

Aufgrund der Tatsache, dass bei dieser Art von CLIL-Unterricht das Leseverständnis wahrscheinlich insgesamt weniger im Fokus und die Gruppe ohne CLIL insgesamt im Bereich «Lesen» etwas besser abschnitt, könnten diese Ergebnisse darauf hinweisen, dass es sich um eine leistungsstarke Kontrollgruppe handelt. Die Kontrollgruppe, mehrheitlich bestehend aus 6. Klassen, konnte bereits beim ersten Erhebungszeitpunkt im Bereich «Lesen» höhere Resultate erzielen als die CLIL-Klassen (vgl. Abbildungen 5-6). Die Ergebnisse zum «Lesen» wurden deshalb als erstes präsentiert, weil sie hilfreiche Einblicke in die komplexe Gruppenstruktur der Teilnehmenden bieten und auch die Schwierigkeit der Vergleichbarkeit der Gruppe mit und ohne CLIL anschaulich aufzeigt.

Ergebnisse: Hören

Auch der Test im Bereich «Hören» umfasste drei alltagsbezogene Hörverständnis-Aufgaben: Im ersten Teil stellten sich zwei Teenager vor (10 Punkte), dann sprachen zwei Personen über ihre Erfahrungen mit dem Erlernen von Skifahren und Snowboarden (5 Punkte) und im dritten Teil ging es um detaillierte Durchsagen am Flughafen (5 Punkte). Maximal konnten 20 Punkte erzielt werden. Die Tabelle 5 zeigt die erreichten Durchschnittspunkte der Lernenden mit CLIL vs. jenen ohne CLIL.

Tabelle 5: Durchschnitt des erreichten Punktetotals «Hören» der Lernenden ohne und mit CLIL bei T2 vs. T1

T2 (Juli 2025)	Total Punkte (max. 20 P.)	T1 (September 2024)	Total Punkte (max. 20 P.)
non-CLIL	15.74	non-CLIL	14.11
CLIL	15.95	CLIL	14.24

Erneut erfreulich zu sehen ist, dass beide Gruppen am Erhebungszeitpunkt 2 (T2) am Ende Schuljahr besser abschnitten als zu Beginn des Schuljahres. Auch zeigen die höheren Werte in der Tabelle links, dass die Schüler:innen mit CLIL beim Hörverständnis mit durchschnittlichen 15.95 Punkten leicht bessere Leistungen zeigten als die Schüler:innen ohne CLIL, die im Durchschnitt 15.74 Punkte erreichten. Der durchgeführte Welch-Test wies jedoch wiederum – wie auch schon bei der ersten Erhebung – keine signifikanten Unterschiede ($t(143.3) = -0.505$, $p = 0.615$) zwischen den Lernenden ohne CLIL ($M = 15.74$; $SD = 3.16$) und jenen mit CLIL ($M = 15.95$; $SD = 2.41$) aus. Der Effekt von $d=0.08$ ist als klein zu beurteilen (Cohen, 1988).

Wie die differenzierte Aufteilung der Ergebnisse nach Untergruppen aussieht, wird in der Abbildung 7 ersichtlich.

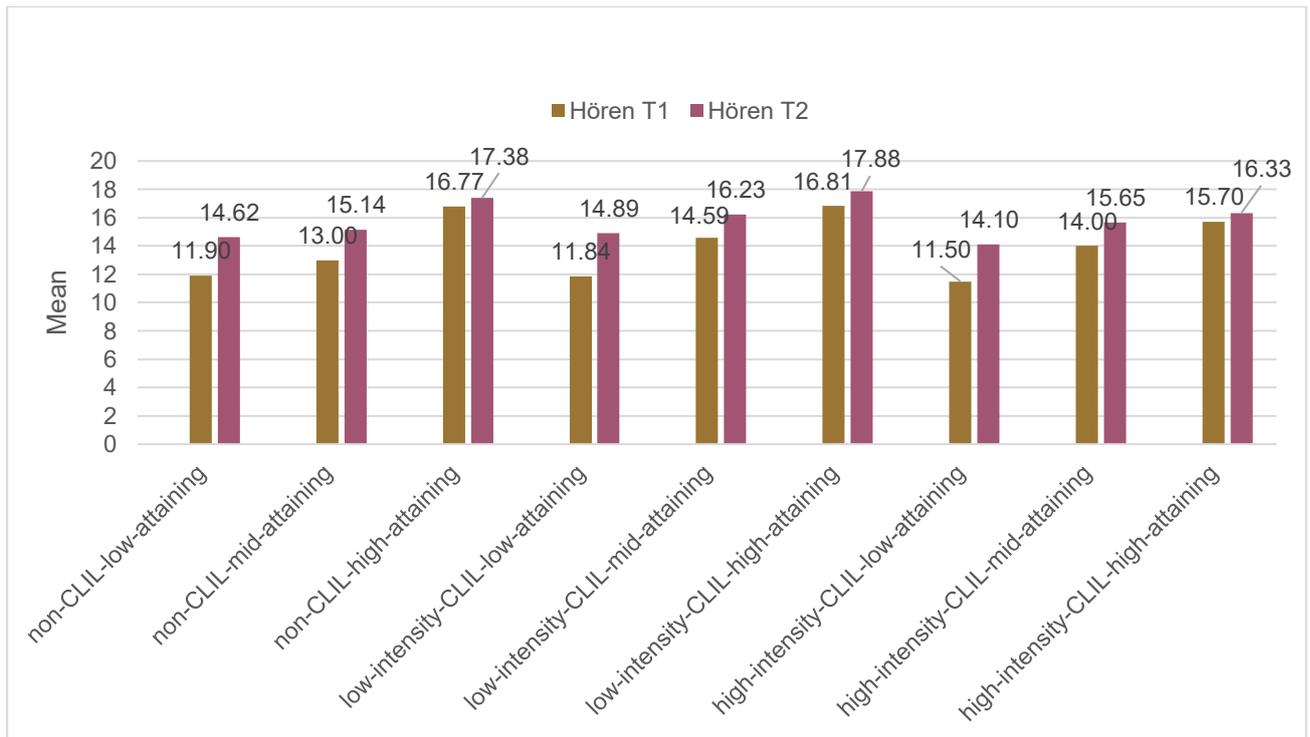


Abbildung 7: Durchschnittliche Ergebnisse im Bereich «Hören» nach Untergruppen und Erhebungszeitpunkten

Auch die gemittelten Ergebnisse aufgeschlüsselt nach Untergruppen visualisiert in Abbildung 7 zeigen, dass jede Untergruppe über das Schuljahr hinweg bessere Resultate erzielen konnte. Die Punkteverteilung fällt dabei kongruent zu den Leistungsgruppen aus, d.h. die stärkeren Lernenden erzielten bessere Resultate als ihre Peers mit tieferen Englischkenntnissen innerhalb der Gruppe mit denselben Kontaktzeiten zu Englisch/CLIL. Auch ist erkennbar, dass die unterschiedlichen Untergruppen an verschiedenen Ausgangspunkten starteten, z.B. die lernstärkeren Gruppen erreichten bereits beim ersten Messzeitpunkt deutlich höhere Punktezahl als die Lernenden mit tieferen Englischkenntnissen. Ebenfalls eine interessante Beobachtung ist, dass die low-attainers in jeder Gruppe innerhalb eines Schuljahres die grössten Fortschritte erzielen konnten.

Wie die Resultate aussehen, wenn nur die vergleichbaren 6. Klassen berücksichtigt werden, ist in der nachfolgenden Abbildung 8 dargestellt.

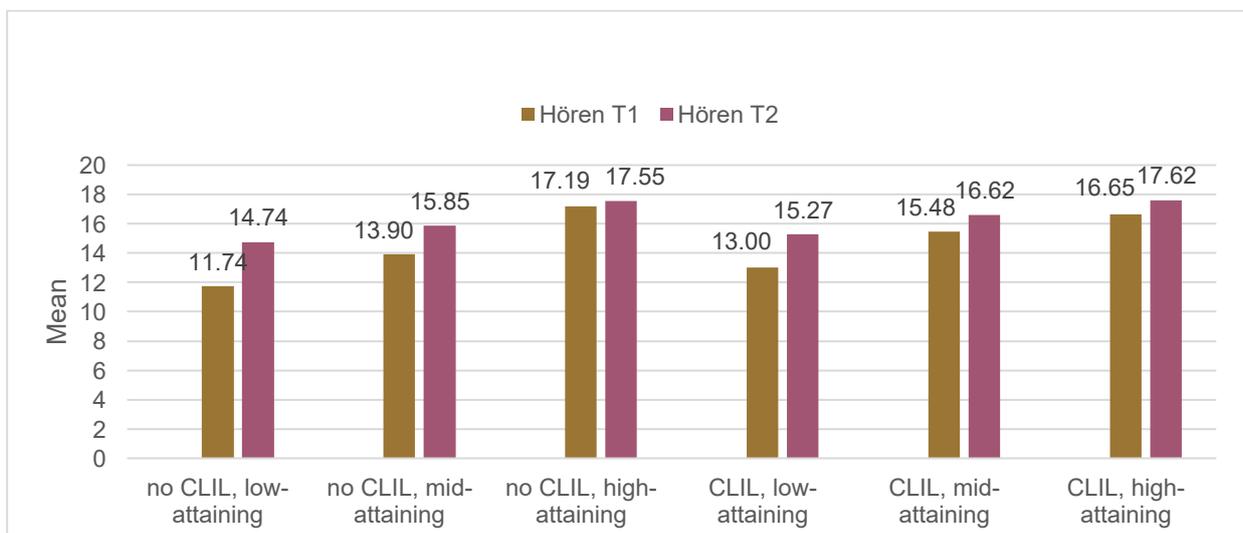


Abbildung 8: Durchschnittliche Ergebnisse der 6. Klassen «Hören» nach Stärkegruppen und Erhebungszeitpunkten

Die CLIL-6.Klasskinder einer jeder Stärkegruppe erreichten am zweiten Messzeitpunkt bessere Resultate als jene Lernende ohne CLIL. Auch hier scheinen die low-attainers beider Gruppen die grössten Fortschritte erzielt zu haben. Die high-attainers mit und ohne CLIL, als auch die mid-attainers mit CLIL – erneut aufgrund der Tatsache, dass sie bereits beim ersten Test erfreulich hoch abschnitten – konnten sich weniger deutlich steigern. Ob es auch statistisch signifikante Unterschiede zwischen den zwei Messzeitpunkten für jeden Gruppe gibt, ist in der nachfolgenden Tabelle zusammengefasst.

Tabelle 6: Statistische Werte zum Test «Hören» der 6. Klasskinder

Untergruppe	Differenz (T1-T2)	Median T1	Median T2	Wilcoxon-Test (exakt)	Effekt (r)
Non-CLIL, low-attainers	3.00	12.50	15.00	$z = -2.499, p = 0.010, n=19$	0.57
Non-CLIL, mid-attainers	1.95	13.75	16.00	$z = -2.533, p = 0.010, n=20$	0.57
Non-CLIL, high-attainers	0.36	17.00	18.00	$z = -1.040, p = 0.313, n=22$	0.22
CLIL, low-attainers	2.27	13.25	15.00	$z = -3.115, p < 0.001, n=22$	0.66
CLIL, mid-attainers	1.14	16.00	16.00	$z = -1.220, p = 0.235, n=21$	0.27
CLIL, high-attainers	0.97	17.25	18.00	$z = -1.722, p = 0.087, n=13$	0.48

Wie in der Tabelle 6 ersichtlich, zeigen sich tatsächlich signifikante Unterschiede und starke Effekte im Bereich «Hören» zwischen den zwei Messzeitpunkten bei den low-attainers mit und ohne CLIL, sowie bei den mid-attainers ohne CLIL. Diese Kinder, die zu Beginn die tiefsten Resultate auswiesen, konnten sich folglich am meisten verbessern.

Zusammengefasst deuten die Ergebnisse zum «Hören» darauf hin, dass die lernschwächeren Kinder mit und ohne CLIL am meisten Fortschritte erzielen konnten. Da dies auch für die Kinder ohne CLIL zutrifft, scheint der fremdsprachliche Unterricht insgesamt diese Kinder besonders gut im Bereich «Hören» zu fördern. Andererseits muss auch angefügt werden, dass die lernstärkeren Kinder sowie die mid-attainers mit CLIL, bereits sehr gute Leistungen bei der ersten Testerhebung zeigten und es deshalb auch nicht mehr so viel Raum nach oben gibt, um sich zu steigern bei diesen Tests mit einem Maximum von 20 Punkten. Dass alle 6. Klasskinder mit CLIL am Ende besser abschnitten als die vergleichbare – wie angenommen – lernstarke Kontrollgruppe könnte schon darauf hinweisen, dass sich CLIL insgesamt positiv auf das Hörverständnis der Lernenden auswirkt.

Wie die Resultate aussehen bei jenen Kompetenzen, die nicht durch strikte Testvorgaben und enge Aufgabenstellungen beeinflusst werden, sondern bei denen die Lernenden die Anwendung ihrer Sprache freier unter Beweis stellen können, wird nachfolgend thematisiert.

Ergebnisse: Sprechen

Aus organisatorischen Gründen wurden die «Speaking Tests» nur mit je drei Kindern mit durchschnittlichen Englischkenntnissen (mid-attainers) aus jeder Klasse durchgeführt (N=36). Die Sprechaufgabe beinhaltet, dass die Lernenden ein fiktives Telefongespräch mit einer Gastfamilie in England durchführten, der sie ihre Ankunftszeit und einen Treffpunkt mitteilen, sowie ihr Aussehen, ihre Kleidung und ihr Gepäckstück beschreiben mussten. Sie konnten dafür vorgängig einige Notizen machen. Ebenfalls wurden sie etwas Weiteres zu ihren Essensvorlieben und Allergien gefragt. Damit wurde ihre spontane Interaktionsfähigkeit geprüft. Die «Speaking Tests» wurden beurteilt in Bezug auf Aufgabenerfüllung (4 Punkte), Umfang des Wortschatzes (3 Punkte), Redefluss (3 Punkte), formale Korrektheit (3 Punkte) und Interaktionsfähigkeit (4 Punkte). Maximal waren somit 18 Punkte möglich. Im Gegensatz zum ersten Test, bei dem sich die Schüler:innen vorstellen mussten, war dieser Test in Bezug auf Wortschatz und Aufgabenkomplexität etwas anspruchsvoller. Dies zeigt sich auch an den insgesamt tieferen Durchschnittswerten dargestellt in Tabelle 7.

Tabelle 7: Durchschnitt der erreichten Punkte «Sprechen» der Lernenden ohne und mit CLIL bei T2 vs. T1

T2 (Juni 2025)	Total Punkte (max. 18 P.)	T1 (September 2024)	Total Punkte (max. 18 P.)
non-CLIL	12.53	non-CLIL	13.44
CLIL	13.76	CLIL	14.06

Wie in der obigen Tabelle links ersichtlich, schnitten die 21 mid-attainers mit CLIL im «Speaking Test» mit durchschnittlich 13.76 Punkten leicht besser ab als die 15 mid-attainers ohne CLIL, die im Durchschnitt 12.53 Punkte erzielten. Während auf deskriptiver Ebene Unterschiede zwischen den zwei Gruppen zu erkennen sind, zeigt der durchgeführte Welch-Test keine signifikanten Unterschiede ($t(21.37) = -1.635, p = 0.117$) zwischen den Lernenden ohne CLIL ($M = 12.53; SD = 2.56$) und jenen mit CLIL ($M = 13.76; SD = 1.58$). Der Effekt von $d=0.60$ ist als mittel einzustufen.

Die nachfolgende Tabelle 8 zeigt die Aufteilung der erreichten Punkte zu jedem der Teilaspekte am zweiten Erhebungszeitpunkt (T2). Die Ergebnisse verdeutlichen, dass Lernende mit CLIL in den Bereichen «Aufgabenerfüllung», «Wortschatz», «Flüssigkeit» und «Interaktion» besser abschnitten. Die Lernenden ohne CLIL hingegen im Bereich «Korrektheit» überlegen waren. Der durchgeführte Welch-Test berichtete signifikante Unterschiede in Bezug auf die Aufgabenerfüllung ($t(25.37) = -3.485, p = 0.002$) und Flüssigkeit ($t(22.20) = -2.826, p = 0.010$). Die Effektstärken von $d=1.23$, resp. $d=1.03$ sind als hoch einzustufen. Das lässt vermuten, dass die CLIL-Lernenden die verlangten Inhalte in der Sprechaufgabe deutlich besser berücksichtigt und flüssiger gesprochen haben.

Tabelle 8: Detaillierte Übersicht aller Durchschnitte im Test «Sprechen» der Lernenden ohne und mit CLIL bei T2

	Aufgabenerfüllung (max. 4 P.)	Wortschatz (max. 3 P.)	Flüssigkeit (max. 3 P.)	Korrektheit (max. 3 P.)	Interaktion (max. 4 P.)
non-CLIL	2.87	2.53	2.20	2.47	2.47
CLIL	3.67	2.62	2.76	2.14	2.57

Schaut man sich die Ergebnisse nicht nur bezüglich CLIL vs. nicht CLIL, sondern auch hinsichtlich der Intensität von CLIL an, so ergibt sich folgendes Bild zu den Gesamtpunkten beim Test «Sprechen»:

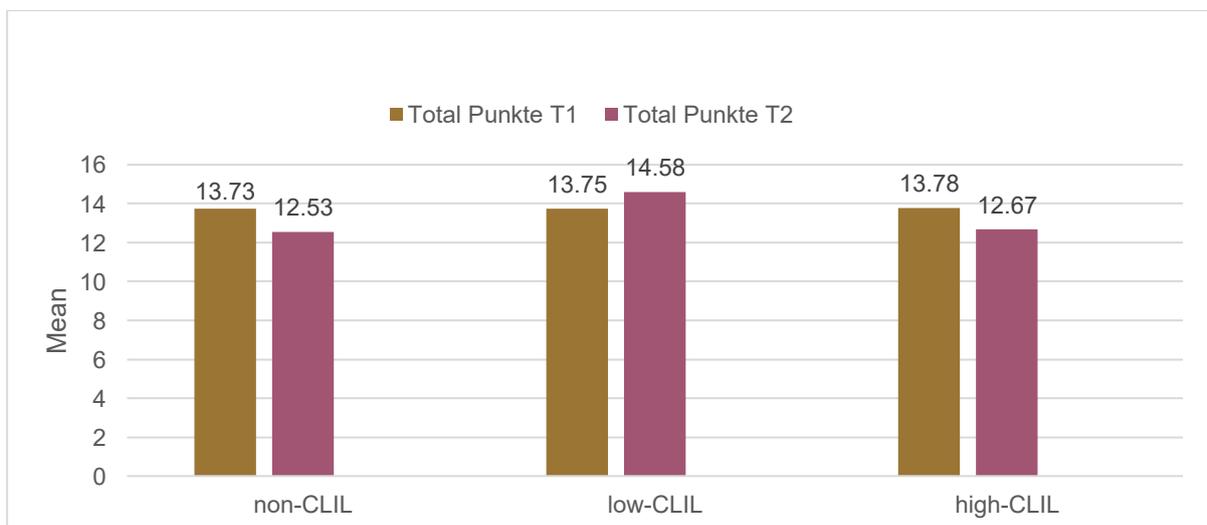


Abbildung 9: Durchschnittliche Ergebnisse im Bereich «Sprechen» nach Untergruppen und Erhebungszeitpunkten

Die Abbildung 9 weist darauf hin, dass die Lernenden ohne CLIL und jene mit hohem Anteil an CLIL-Lektionen, leicht tiefere Leistungen bei der Sprechaufgabe an der zweiten Testdurchführung (T2) zeigten, während sich die Schüler:innen mit wenigen CLIL-Lektionen verbessern konnten.

Die Resultate wurden ebenfalls statistisch analysiert und deren Ergebnisse sind in der nachfolgenden Tabelle 9 zusammengefasst.

Tabelle 9: Statistische Werte zum Test «Sprechen»

Untergruppe	Differenz Mittelwert T1-T2	Median T1	Median T2	Wilcoxon-Test (exakt)	Effekt (r)
non-CLIL	-1.2	13.00	13.00	$z = -1.431, p = 0.152, n=15$	0.36
low-CLIL	0.83	14.00	15.00	$z = -1.996, p = 0.046, n=12$	0.58
high-CLIL	1.11	14.00	12.00	$z = -1.445, p = 0.149, n=9$	0.48

Nur bei der Gruppe mit wenig CLIL-Lektionen zeigen sich signifikante Unterschiede zwischen den beiden Messzeitpunkten. Der Effekt (r)³ mit 0.58 ist als stark einzuschätzen (Cohen, 1992).

Zusammengefasst kann Folgendes zum Bereich «Sprechen» festgehalten werden: Betrachtet man die CLIL-Gruppen insgesamt (vgl. Tabellen 7 und 8), so scheinen diese Lernenden in den Teilbereichen «Aufgabenerfüllung», «Wortschatz», «Redefluss» und «Interaktion» den Lernenden ohne CLIL überlegen zu sein. Die CLIL-Lernenden schnitten in den Teilbereichen alle verlangten Informationen mitzuteilen und sich flüssend auszudrücken gar signifikant besser ab als ihre Peers ohne CLIL. Dies kann allenfalls damit erklärt werden, dass es in den CLIL-Lektionen vorwiegend darum geht sich mündlich gewandt und inhaltsorientiert auszudrücken und mit anderen zu interagieren. Auf diese Weise erweitert sich auch der Wortschatz. Die positiven Ergebnisse könnten darauf hinweisen, dass die CLIL-Lektionen sich vorteilhaft für die Schüler:innen in diesen Teilbereichen des Sprechens auswirkten. Vorteile machten sich jedoch nicht im Bereich «Korrektheit» bemerkbar – da zeigten die Schüler:innen ohne CLIL bessere Leistungen. Das formale, richtige Sprechen steht bei den CLIL-Lektionen wahrscheinlich weniger im Vordergrund. Jedoch sollte sich CLIL auch nicht nachteilig auf das korrekte Sprechen auswirken. Die positiven Resultate im eher analytischen Bereich Korrektheit der Lernenden ohne CLIL könnten deshalb erneut dahingehend interpretiert werden, dass es sich bei der Kontrollgruppe um eine leistungsstarke Gruppe handelt, die sich in diesem Bereich auch ohne zusätzliche CLIL-Lektionen behaupten konnte. Dass die Kontrollgruppe ohne CLIL leistungsstark ist, bestätigt sich auch beim Betrachten der differenzierten Resultate aufgeteilt nach Intensitätsstufe (vgl. Abbildung 9). Die Lernenden ohne CLIL zeigten fast identische Leistungen wie jene mit viel CLIL. Insgesamt könnte das auch zeigen, dass die erhöhte Schwierigkeit des Tests «Sprechen» am zweiten Erhebungszeitpunkt den Lernenden ohne CLIL (mehrheitlich 6. Klasskinder) und jene mit höherem CLIL-Anteil (mehrheitlich 5. Klasskinder) etwas zugesetzt hat. Hingegen konnte sich die Gruppe, die bislang weniger lange und weniger intensiv CLIL-Lektionen hatte, beim Sprechen signifikante Fortschritte erzielen. Dies kann allenfalls damit erklärt werden, dass einerseits der Novelty-Effekt und die damit verbundene erhöhte Motivation diese Lernenden anspornte; andererseits, dass diese Gruppe mehrheitlich aus 6. Klasskindern besteht und es anzunehmen ist, dass ihr zusätzliches Lernjahr bei dieser Aufgabe einen positiven Einfluss hatte.

Ergebnisse: Schreiben

Im Bereich «Schreiben» verfassten die Schüler:innen eine Nachricht, in der sie ihr Zimmer und dessen Einrichtung beschrieben. Dies beinhaltete, dass sie auf Englisch die Einrichtungsgegenstände sowie deren Farben, Grössen und Anordnung im Raum beschrieben. Die verfassten Texte wurden mittels Kriterienraster beurteilt in Bezug auf Aufgabenerfüllung (4 Punkte), Umfang des Wortschatzes (4 Punkte), Verwendung der Grammatik und Strukturen (4 Punkte) und formale Korrektheit (Rechtschreibung) (4 Punkte). Maximal waren 16 Punkte möglich.

Die Tabelle 10 zeigt die Durchschnittswerte der erreichten Punkte der Gruppe ohne CLIL im Vergleich zur Gruppe mit CLIL. Aufgrund der tieferen Gesamtpunkte bei der zweiten Messdurchführung ist anzunehmen, dass die Schreibaufgabe am Ende des Schuljahres etwas anspruchsvoller war für alle Schüler:innen.

³ Gemäss Cohen (1992) bedeutet $r < 0.3$ ein schwacher Effekt, $0.3 < r < 0.5$ einen mittleren Effekt, $r > 0.5$ ein starker Effekt.

Tabelle 10: Durchschnitt des erreichten Punktetotals im Test «Schreiben» der Lernenden ohne und mit CLIL bei T2 vs. T1

T2 (Juni 2025)	Total Punkte (max. 16 P.)	T1 (September 2024)	Total Punkte (max. 16 P.)
non-CLIL	9.83	non-CLIL	10.80
CLIL	10.03	CLIL	10.74

Wie in der Tabelle links ersichtlich, schnitten die Schüler:innen mit CLIL bei der Aufgabe «Schreiben» mit durchschnittlich 10.03 Punkten Ende Schuljahr leicht besser ab als Schüler:innen ohne CLIL. Die Unterschiede sind jedoch marginal, weshalb es wenig überraschend ist, dass der durchgeführte Welch-Test keine signifikanten Unterschiede ($t(186) = 0.56, p = 0.58$) zwischen den Lernenden mit CLIL ($M = 10.03; SD = 2.40$) und jenen ohne CLIL ($M = 9.38; SD = 2.32$) berichtete. Die Effektstärke $d=0.083$ ist als klein zu beurteilen.

Die nachfolgende Tabelle 11 fasst die erreichten Punkte zu jedem Teilbereich im Test «Schreiben» zusammen.

Tabelle 11: Detaillierte Übersicht aller Durchschnitte im Test «Schreiben» der Lernenden ohne und mit CLIL am T2

	Aufgabenerfüllung (max. 4 P.)	Wortschatz (max. 4 P.)	Grammatik (max. 4 P.)	Korrektheit (max. 4 P.)
non-CLIL	2.93	2.61	2.21	2.29
CLIL	2.83	2.68	2.61	2.26

Die Tabelle verdeutlicht, dass Lernende ohne CLIL in den Bereichen «Aufgabenerfüllung» und «Korrektheit» besser abschnitten, die CLIL-Lernenden hingegen im Bereich «Wortschatz» und «Grammatik» überlegen waren. Der durchgeführte Welch-Test zeigt signifikante Unterschiede zwischen den zwei Gruppen in Bezug auf «Grammatik» ($t(162) = 3.90, p < 0.001$). Die Effektstärke mit $d = 0.690$ ist als mittel zu bezeichnen.

Welche durchschnittlichen Ergebnisse die unterschiedlichen Untergruppen im Bereich «Schreiben» insgesamt erzielten, wird in der nachfolgenden Abbildung 10 dargestellt.

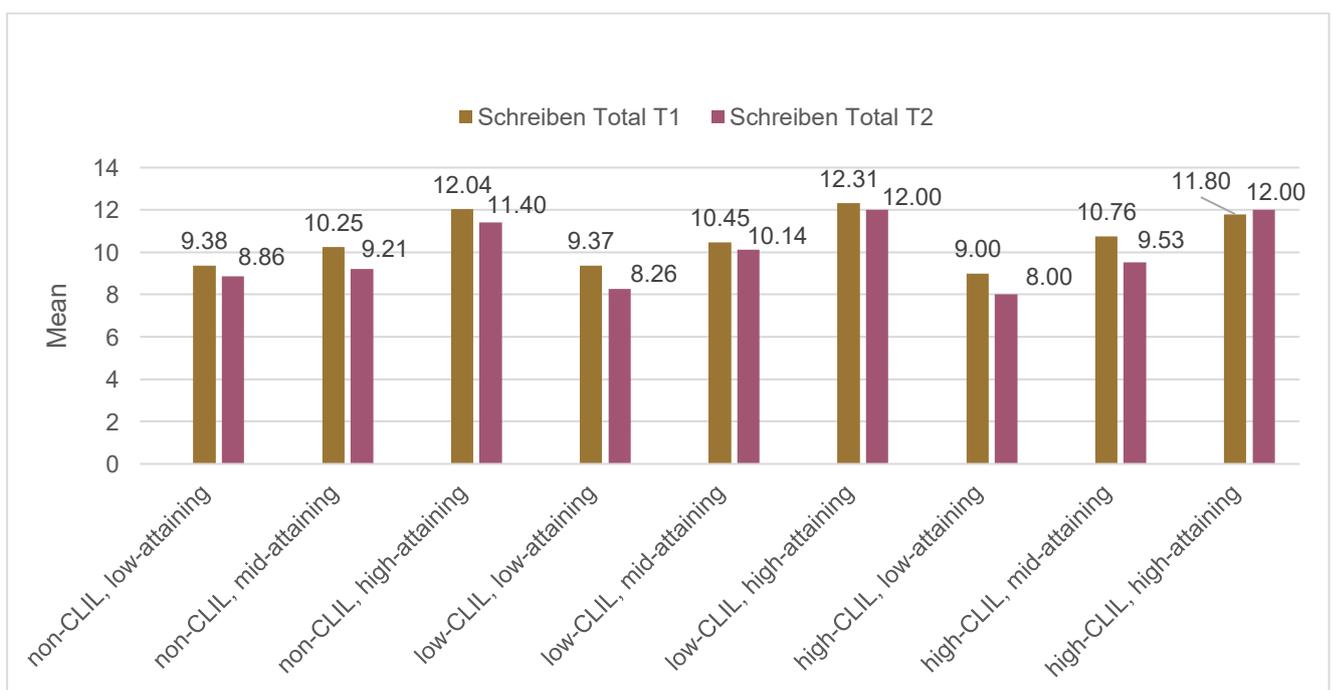


Abbildung 10: Durchschnittliche Ergebnisse des Tests «Schreiben» nach Untergruppen und Erhebungszeitpunkten

Wie bereits angetönt, machten die erhöhten Anforderungen der Schreibaufgabe allen Untergruppen ausser den high-attainers mit hohem CLIL-Anteil zu schaffen – dies ist an den tieferen Mittelwerten beim Messzeitpunkt Ende des Schuljahres (T2) zu erkennen. Insgesamt ist ebenfalls zu erkennen, dass alle high-attainers weniger grosse Punkteunterschiede verzeichneten als die anderen Leistungsgruppen.

Wie diese Verteilung aussieht, wenn man nur die 6. Klasskinder vergleicht, visualisiert die nachfolgende Abbildung 11.

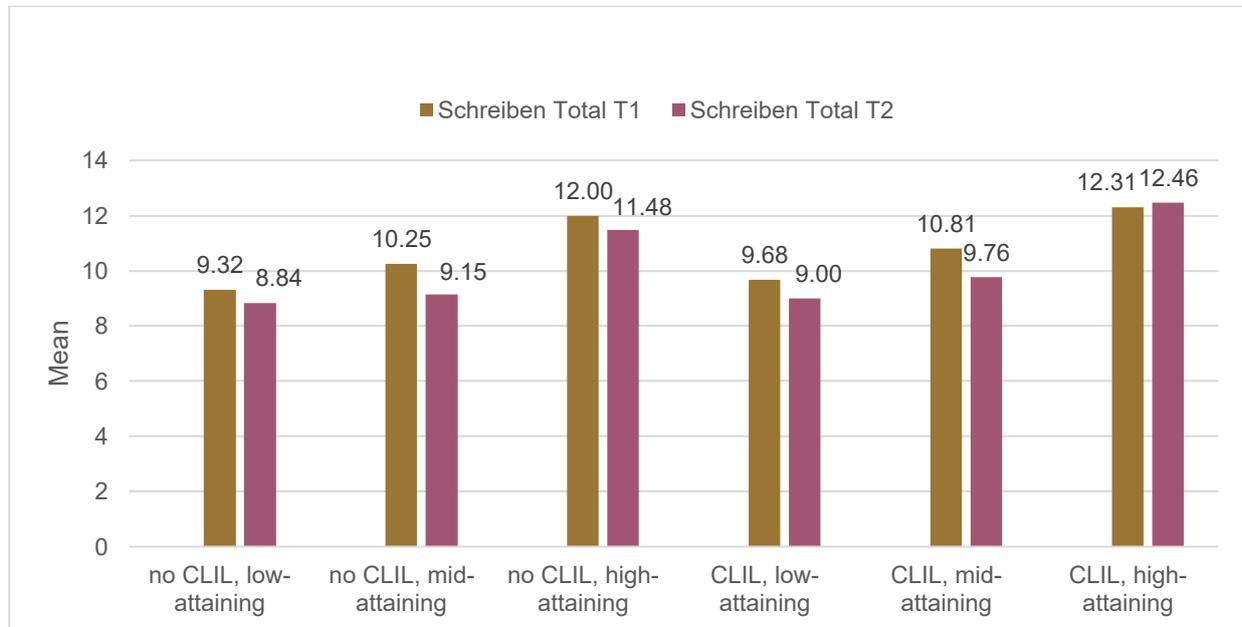


Abbildung 11: Durchschnittliche Ergebnisse der 6. Klassen «Schreiben» nach Stärkegruppen und Erhebungszeitpunkten

Auch diese Grafik verdeutlicht, dass sich einzig die lernstarken 6. Klasskinder mit CLIL beim Schreiben steigern konnten. Bei allen anderen Untergruppen liegen die Resultate beim Messzeitpunkt 2 tiefer als beim Messzeitpunkt 1. Die Unterschiede scheinen jedoch weniger hoch zu sein als bei den oben visualisierten Durchschnitts, die auch die 5. Klasskinder mitbeinhalteten (vgl. Abbildung 10). Ebenfalls ist ersichtlich, dass die drei Untergruppen der 6. Klasskinder mit CLIL im Vergleich zur gleichen Leistungsgruppe besser abschnitten – das ist ein erfreuliches Ergebnis – wohlwissend, dass es sich um eine starke Kontrollgruppe handelt. Ob sich die Ergebnisse der 6. Klasskinder signifikant innerhalb eines Schuljahres veränderten, zeigt die nachfolgende Tabelle 12.

Tabelle 12: Statistische Werte zum Test «Schreiben» der 6. Klasskinder

Untergruppe	Differenz (T1-T2)	Median T1	Median T2	Wilcoxon-Test (exakt)	Effekt (r)
Non-CLIL, low-attainers	-0.48	9.50	9.00	$z = -1.555, p = 0.130, n=19$	0.36
Non-CLIL, mid-attainers	-1.10	11.00	9.00	$z = -2.913, p = 0.003, n=20$	0.65
Non-CLIL, high-attainers	-0.52	12.00	11.00	$z = -1.129, p = 0.281, n=21$	0.25
CLIL, low-attainers	-0.68	10.00	9.00	$z = -2.016, p = 0.044, n=22$	0.43
CLIL, mid-attainers	-1.05	11.00	10.00	$z = -2.604, p = 0.007, n=21$	0.57
CLIL, high-attainers	0.15	12.00	12.50	$z = -0.462, p = 0.796, n=13$	0.13

Die Resultate der mid-attainers aus beiden Gruppen veränderten sich signifikant – doch leider nachteilig gegenüber dem ersten Messzeitpunkt. Auch die Effektgrössen dieser beiden Gruppen sind als hoch zu beurteilen. Die Ergebnisse im Bereich «Schreiben» der low-attainers mit CLIL veränderten sich ebenfalls auf signifikante Weise, die Effektstärke sind als mittel einzustufen. Die hier berechneten Unterschiede sind jedoch eher ein Indiz, dass die Aufgabe tatsächlich anspruchsvoll war zum Schuljahresschluss.

Zusammengefasst kann zum «Schreiben» festgehalten werden, dass die CLIL-Lernenden die Schreibaufgaben insgesamt und bezogen auf die Teilbereiche «Wortschatz» und «Grammatik» bessere Resultate erzielten als die Kinder ohne CLIL. Im Bereich der Verwendung der richtigen Satzstrukturen (Grammatik) waren ihre Ergebnisse als Gesamtgruppe sogar signifikant besser. Die Schreibaufgabe am Ende des Schuljahres war anspruchsvoller als jene zu Beginn des Schuljahres wie die oben dargestellten Ergebnisse mit rückläufigen Punktezahlen bei den meisten Untergruppen zeigen. Dass sich die Ergebnisse der durchschnittlichen 6. Klasskinder mit und ohne CLIL auf signifikante Weise veränderten, könnte daran liegen, dass diese Kinder bei der ersten Aufgabe recht gut abschnitten und nun bei der anspruchsvolleren Aufgabe keine vergleichbaren Leistungen zeigen konnten. Einzig die high-attainers mit CLIL, also die lernstärksten Kinder mit den meisten CLIL-Lektionen, konnte diese anspruchsvolle Aufgabe Ende Schuljahr besser lösen als die erste Schreibaufgabe. Diese Beobachtung könnte darauf hinweisen, dass die zusätzliche Kontaktzeit mit Englisch dank CLIL vorteilhaft sein kann – sich jedoch einzig die Lernenden mit den besten fremdsprachlichen Ausgangsbedingungen diesen Vorteil tatsächlich nutzen konnten, bei allen anderen dämpfte die schwierige Aufgabenstellung die Verbesserungsleistungen.

Ergebnisse: Problemlöse Verhalten

Wie bereits angesprochen, lösten alle Kinder auch eine Reihe von Problemlöseaufgaben namens «Tower of London», um ihre Planungsfähigkeit – welche als einen wichtigen Teil von Problemlöseverhalten angesehen wird – unter Beweis zu stellen.

Die Tabelle 13 zeigt die Durchschnittswerte der gesamt erreichten Punkte in der Aufgabenserie «Tower of London» der Gruppe ohne im Vergleich zur Gruppe mit CLIL. Das Manual des «Tower of London» legt nahe, ein Prozentrang von 16 bis 24 als unterdurchschnittlich bis durchschnittlich, solche von 25 bis 75 als durchschnittlich, solche zwischen 76 und 84 als durchschnittlich bis überdurchschnittlich und ein Prozentrang grösser 84 als klar überdurchschnittlich zu interpretieren (Kaller et al., 2011).

Tabelle 13: Durchschnitt der Prozenträge beim «Tower of London» der Lernenden ohne und mit CLIL bei T2 vs. T1

T2 (Juni 2025)	Prozentrang (max. 100)	T1 (September 2024)	Prozentrang (max. 100)
non-CLIL	35.46	non-CLIL	17.91
CLIL	36.58	CLIL	26.52

Wie in der linken Tabelle ersichtlich, liegen die Resultate beider Gruppen beim Messzeitpunkt zwei im durchschnittlichen Bereich. Dies im starken Kontrast zum letzten Messzeitpunkt, wo die Resultate zwischen den zwei Gruppen CLIL und non-CLIL gar signifikant unterschiedlich ausfielen. Der erneut durchgeführte Welch-Test weist dieses Mal keine signifikanten Unterschiede ($t(182) = -0.310, p = .757$) zwischen den Lernenden ohne CLIL ($M = 35.46; SD = 22.60$) und jenen mit CLIL ($M = 36.58; SD = 25.33$) aus. Die Effektstärke von 0.046 ist als klein zu beurteilen.

Wie die Abbildung 12 aufgezeigt, konnten praktisch alle Untergruppen – ausser die low- und high-attainers mit viel CLIL – sehr grosse Fortschritte bei diesem Test erzielen. Das kann einerseits damit zusammenhängen, dass die Kinder den fast identischen Test beim zweiten Messzeitpunkt bereits einmal durchgeführt hatten und sie nun das Vorgehen und den Ablauf besser verstanden hatten. Andererseits kann auch das zusätzliche Schuljahr die Fähigkeit Probleme zu lösen positiv beeinflusst haben.

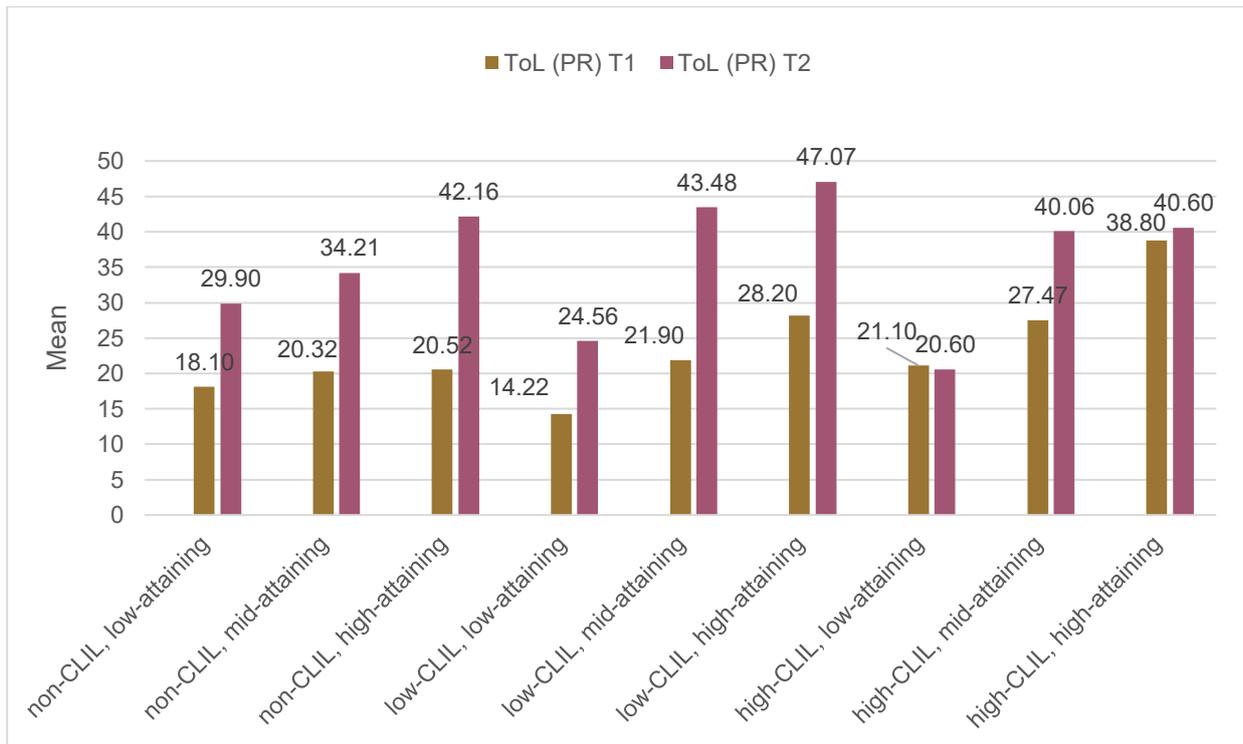


Abbildung 12: Durchschnittliche Ergebnisse des Tests «Tower of London» nach Untergruppen und Erhebungszeitpunkten

Diese starken Veränderungen zwischen den zwei Testzeitpunkten sind auch weiterhin ersichtlich, wenn nur die 6. Klasskinder miteinander verglichen werden. Sie scheinen noch verstärkt zu werden, zum Beispiel bei den high-attainers mit CLIL. Dies verdeutlicht die Abbildung 13.

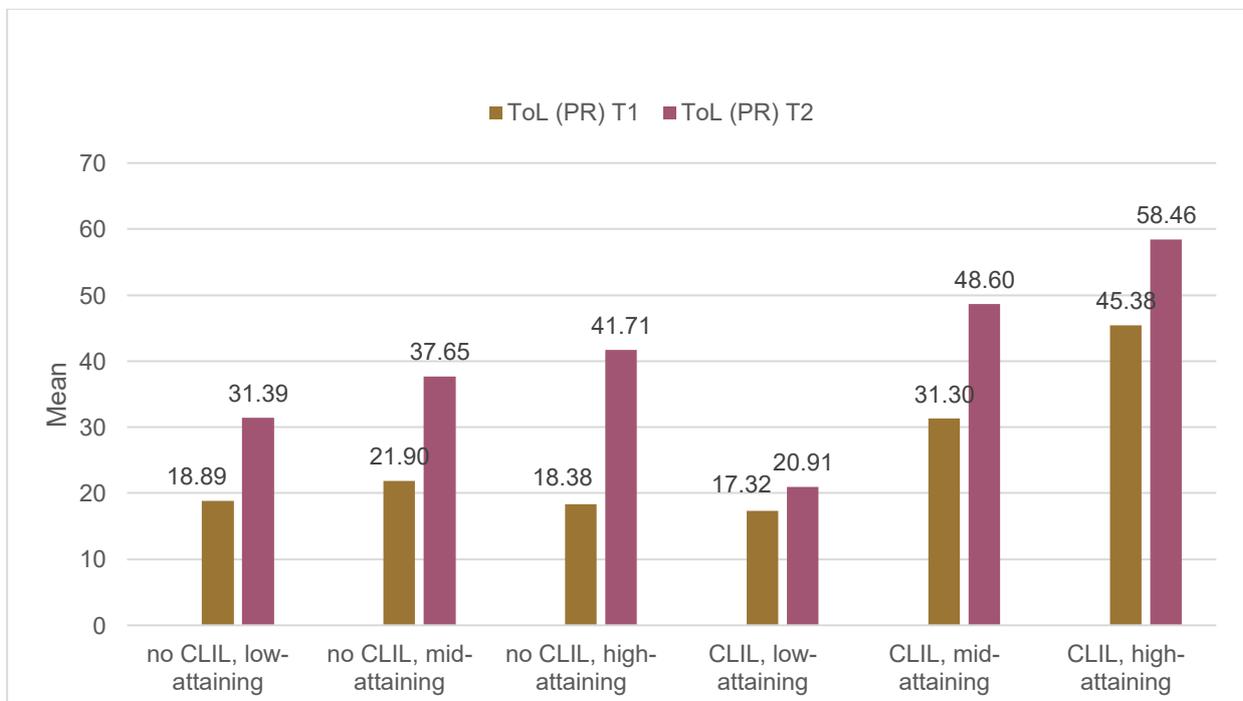


Abbildung 13: Durchschnittliche Ergebnisse der 6. Klassen zum Test «Tower of London» nach Stärkegruppen und Erhebungszeitpunkten

Die Ergebnisse wurden ebenfalls statistisch geprüft. Wie in der Tabelle 14 ersichtlich konnten sich die mid- und high-attainers ohne CLIL signifikant steigern. Diese lernstarken Kinder, ungeachtet von CLIL, schnitten am ersten Erhebungszeitpunkt eher tief ab. Wahrscheinlich half ihnen die erneute Testdurchführung sich signifikant zu steigern. Ebenfalls die mid-attainers mit CLIL konnten sich deutlich steigern. Auch hier zeigen sich Effektstärken im hohen Bereich. Da die mid- und high-attainers mit CLIL bereits am ersten Erhebungszeitpunkt deutlich besser abschnitten als die Peers ohne CLIL, könnte ein Indiz sein, dass CLIL im Bereich des Problemlöseverhaltens für lernstarke Kinder einen Vorteil ausweist. Weniger jedoch für lernschwächere Kinder. Diese – ungeachtet von CLIL – schnitten bei beiden Testerhebungen eher tief ab und konnten sich auch nicht signifikant steigern.

Tabelle 14: Statistische Werte zur Problemlöseaufgabe der 6. Klasskinder

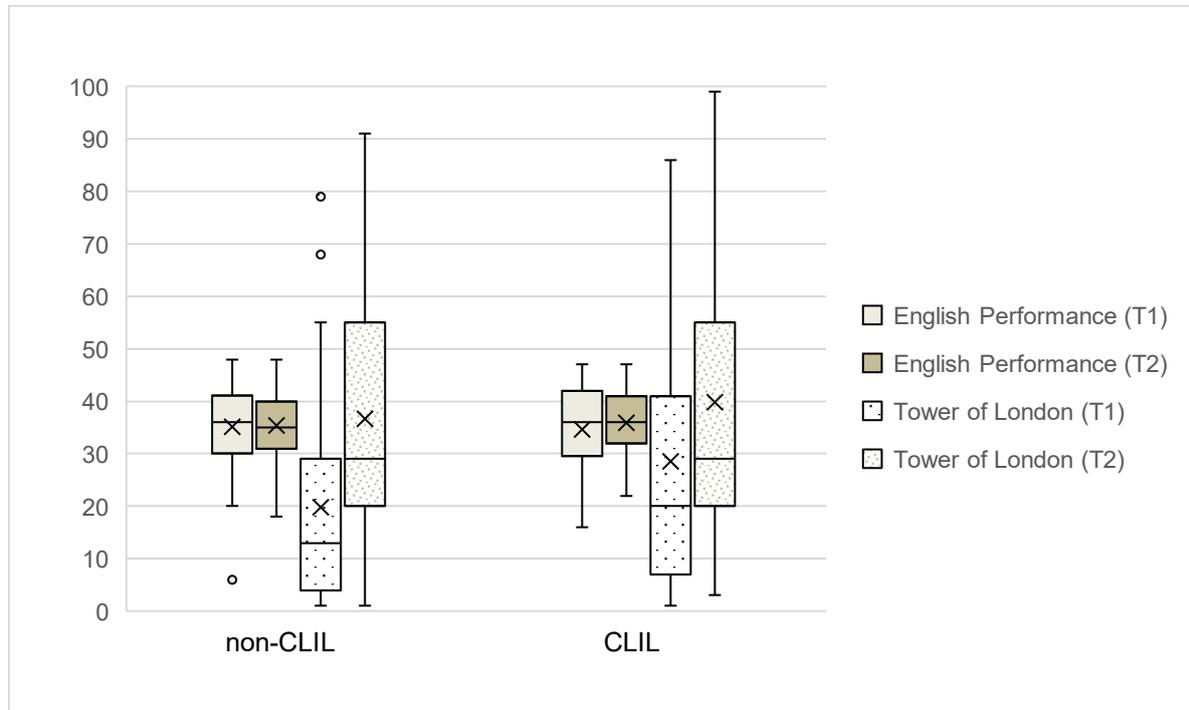
Untergruppe	Differenz (T1-T2)	Median T1	Median T2	Wilcoxon-Test (exakt)	Effekt (r)
Non-CLIL, low-attainers	12.50	13.00	29.00	$z = -1.701, p = 0.091, n=18$	0.40
Non-CLIL, mid-attainers	15.75	13.00	29.00	$z = -2.334, p = 0.018, n=20$	0.52
Non-CLIL, high-attainers	23.33	20.00	41.00	$z = -3.203, p < 0.001, n=21$	0.70
CLIL, low-attainers	3.59	13.00	20.00	$z = -0.486, p = 0.641, n=22$	0.10
CLIL, mid-attainers	17.30	24.50	41.00	$z = -2.354, p = 0.016, n=20$	0.53
CLIL, high-attainers	13.08	48.00	55.00	$z = -1.727, p = 0.084, n=13$	0.48

Zusammengefasst kann an dieser Stelle festgehalten werden, dass beim Problemlöse Test «Tower of London» anders als bei den anderen fremdsprachenbezogenen Tests die mid- und high-attainers die grösseren Fortschritte erzielen konnten. Auch wenn die high-attainers mit CLIL keine signifikanten Fortschritte ausweisen konnten, so zeigen deren Effektstärken einen hohen mittleren Effekt. Ebenfalls auffällig ist, dass der Median der high-attainers mit CLIL bereits beim ersten Testen anfangs Schuljahr sehr hoch ausfiel. Deshalb konnten sie bei der zweiten Testdurchführung wahrscheinlich auch weniger Fortschritte erzielen. Ebenfalls spricht für CLIL, dass die Testresultate der mid-attainers mit CLIL ähnlich hoch ausfielen wie jene der high-attainers ohne CLIL. Das könnte darauf hinweisen, dass die durchschnittlichen Lernenden dank CLIL ihre Problemlösefähigkeiten weiterentwickeln konnten, wie es sonst nur den lernstarken Kindern vorenthalten ist. Auffällig ist ebenfalls, dass sich keine der Gruppe an low-attainers signifikant steigern konnte. Auch wenn die Einteilung in die Gruppe der low-attainers in Bezug auf das Fach Englisch vorgenommen wurde, so könnten diese Resultate dahingehend interpretiert werden, dass diese Untergruppe generell schwächere Dispositionen mitbringt und sie deshalb in Bezug auf Problemlösefähigkeiten auch weniger weitere Ressourcen mobilisieren konnte.

Fortschritte Englishtests und «Tower of London» insgesamt

Die vorherigen, detaillierten Ergebnispräsentationen lassen allenfalls den Blick auf das Wesentliche verlieren. Daher werden abschliessend an dieser Stelle die Ergebnisse für die 6. Klassgruppe als ganze, ohne Untergruppen und alle Resultate zum Englisch aufsummiert betrachtet. Das bedeutet, dass die erreichten Punkte aller drei Englishtests zu den Kompetenzbereichen Lesen, Hören und Schreiben addiert wurden. Diese gebündelten Ergebnisse fassen die «English performance» insgesamt zusammen. Betrachtet man nun diese zusammengefassten Resultate fürs Englisch und den Problemlösetest «Tower of London» zu Beginn (T1) und am Ende des Schuljahres (T2) für die vergleichbaren 6. Klasskinder mit CLIL und ohne CLIL, zeigen sich folgende Veränderungen innerhalb eines Schuljahres (vgl. Abbildung 14).

Abbildung 14: Ergebnisse der 6. Klasskinder mit und ohne CLIL für Englisch insgesamt und Problemlöseverhalten an beiden Erhebungszeitpunkten



Wie die jeweils höheren Mittelwerte in den Boxplots (X) bei T2 zeigen, haben beide Gruppen im Laufe der Zeit bessere Leistungen in Englisch und beim «Tower of London» Test erzielt. In Bezug auf ihre Leistungen in Englisch zeigt das Boxplot auf den ersten Blick keine auffälligen Unterschiede zwischen den beiden Testdurchführungen – insbesondere im Hinblick auf die non-CLIL Schüler:innen. Die statistischen Analysen mittels Wilcoxon-Test zeigen jedoch, dass die CLIL-Schüler:innen im Verlaufe des Schuljahres signifikante Fortschritte erzielt haben ($z = -2,832$, $p = 0,005$, $n = 56$). Die Effektstärke von 0,38 deutet gemäss Cohen auf einen mittleren Effekt hin. Bei den Schüler:innen ohne CLIL gibt es keine signifikanten Unterschiede ($z = -1,489$, $p < 0,069$, $n = 61$). Die Entwicklung der Problemlösefähigkeiten beider Gruppen, gemessen mit dem «Tower of London» Test, ist hingegen in obiger Abbildung deutlich sichtbar. Auch zeigt sich, dass die CLIL-Schüler:innen zu Beginn des Schuljahres deutlich besser abschnitten als die non-CLIL Schüler:innen. Diese Beobachtungen wurden auch durch weitere statistische Analysen bestätigt: Die Unterschiede im Laufe der Zeit für die non-CLIL Gruppe ($z = -4,218$, $p < 0,001$, $n = 59$) und die Schüler:innen mit CLIL ($z = -2,832$, $p = 0,002$, $n = 55$) sind signifikant. Die Effektgrößen sind hoch ($r = 0,55$) für die non-CLIL Gruppe und mittel ($r = 0,38$) für die CLIL-Gruppe.

Zusammenfassung der Ergebnisse

Im Folgenden werden die fünf zentralen Fragestellungen beantwortet und die Ergebnisse zusammenfassend interpretiert.

- 1) Welche Einstellungen haben die Lernenden gegenüber dem Englisch und wie schätzen sie sich ein? Welche motivationalen Einstellungen gegenüber CLIL bringen die CLIL-Schüler:innen mit?

Die Selbsteinschätzungen zum Lernen von Englisch allgemein und CLIL sind äusserst positiv. Auch zeigt die Abbildung 1, dass die meisten Lernenden ein gutes Selbstbild in Bezug auf ihre Englischleistung haben. Aus den Selbsteinschätzungen kommt hervor, dass es tendenziell einen Zusammenhang gibt zwischen Leistungsvoraussetzungen und CLIL-Intensität. Das bedeutet, dass je besser die Lernenden im Englisch sind und je mehr sie damit in Kontakt kommen, desto lieber lernen sie Englisch im Vergleich mit den Lernenden derselben Leistungsstärke (vgl. Abbildung 2). Alle CLIL-Schüler:innen, ausser die low-attainers mit wenig CLIL, stimmten dem Statement «Ich finde es gut, dass wir CLIL haben» vollumfänglich oder eher zu – das zeigt sich an den Mittelwerten, die über dem Wert 3 (= «eher ja») liegen (vgl. Abbildung 3).

- 2) Wie schnitten die CLIL-Lernenden im Vergleich zu jenen Kindern ohne CLIL bei den unterschiedlichen Tests ab?

Die Ergebnisse sprechen auf deskriptiver Ebene zugunsten der Lernenden mit CLIL-Unterricht in den Bereichen «Sprechen», «Hören», «Schreiben» und «Problemlösefähigkeiten». Das sind erfreuliche Befunde! Die Unterschiede erwiesen sich jedoch nicht als signifikant. Einzig bei den Teilbereichen Aufgabenerfüllung und Flüssigkeit bei den Testaufgaben «Sprechen» sowie bei Grammatik im Bereich «Schreiben» schnitten die CLIL-Kinder signifikant besser ab. Dies kann damit erklärt werden, dass der mündlich geführte CLIL-Unterricht den Redefluss und das inhaltliche Kommunizieren positiv beeinflusste. In Bezug auf den Aspekt Grammatik im Bereich «Schreiben» könnten die deutlich besseren Werte dahingehend gedeutet werden, dass die Lernenden korrekte Satzstrukturen im Unterricht (z.B. durch Hören) aufnehmen und beim Schreiben wieder verwenden können. Beim «Lesen» waren die Lernenden ohne CLIL leicht überlegen.

Da die beiden Gruppen CLIL vs. kein CLIL sehr heterogen in Bezug auf Klassenstufe, Intensität und Anzahl Kinder zusammengesetzt sind, müssen diese Ergebnisse mit Vorsicht interpretiert werden. Aussagekräftiger sind deshalb die Resultate im Zusammenhang mit der nächsten Fragestellung.

- 3) Inwiefern haben sich die Ergebnisse zwischen anfangs und Ende des Schuljahres für jede Untergruppe verändert, d.h. unter Berücksichtigung von Englischleistungen und Kontaktzeit mit Englisch/CLIL?

Es ist erfreulich, dass sich jede Untergruppe in den rezeptiven Bereichen «Lesen» und «Hören» über das Schuljahr hinweg steigern konnte. Insbesondere die leistungsschwächeren und durchschnittlichen Untergruppen konnten die grössten Fortschritte erzielen. Auch wenn die Lernenden mit CLIL beim Hören noch deutlichere Fortschritte erzielten als jene ohne CLIL, so konnten sich auch jene ohne CLIL signifikant steigern. Dies kann dahingehend interpretiert werden, dass der fremdsprachliche Unterricht einen besonders positiven Einfluss auf diese Leistungsgruppen hatte, respektive bei diesen vorgegebenen Tests die lernstarken Schüler:innen bereits beim ersten Erhebungszeitpunkt sehr gut abschnitten und sich deshalb wohl nicht mehr allzu stark steigern konnten.

Im Bereich «Sprechen» zeigte sich, dass sich die Lernenden mit wenig CLIL signifikant steigern konnten. Das könnte dahingehend interpretiert werden, dass insbesondere zu Beginn CLIL den Lernenden einen Boost erteilt. Da der CLIL-Unterricht in den handlungsorientierten Fächern mehrheitlich mündlich verläuft, könnte sich diese neue Unterrichtsform positiv auf die Sprechkompetenz der Schüler:innen auswirken. Beide anderen Gruppen zeigten beim Testzeitpunkt Ende des Schuljahres im Bereich «Sprechen» tiefere Leistungen – dies hängt mit der anspruchsvolleren Testaufgabe zusammen. Weil zu erwarten wäre, dass insbesondere die Gruppe mit mehr CLIL und folglich hoher Englischkontaktzeit mindestens ebenso gut abschneiden müsste, lässt vermuten, dass diese Gruppe bestehend aus neun Kindern am Ende Schuljahr weniger Motivation für diese Tests aufbringen konnte. Ähnlich wie beim Bereich «Sprechen» schien der Test «Schreiben» am Ende des Schuljahres schwieriger gewesen zu sein. Dies zeigt sich an den tendenziell tieferen Punkten an der zweiten Erhebung. Einzig die lernstarken Schüler:innen mit viel CLIL konnten bessere Resultate erzielen. Das könnte darauf hinweisen, dass ein hoher CLIL-Anteil bei komplexen Aufgaben ein Vorteil ist für jene Schüler:innen, die diese zusätzliche Kontaktzeit auch produktiv im Sprachgebrauch einsetzen können.

Letztlich wurde erneut die Fähigkeiten Probleme zu lösen mittels des Tests «Tower of London» für jede Untergruppe ausgewertet. Beim zweiten Messzeitpunkt konnten sich die Lernenden ohne CLIL und jene mit wenig CLIL deutlich steigern. Konkret erzielten die durchschnittlichen und starken Lernenden mit wenig und jene ohne CLIL besonders grosse Fortschritte. Dabei waren die Kinder mit CLIL, jenen ohne CLIL etwas überlegen. Ebenfalls ist zu beobachten, dass anders als bei den fremdsprachlichen Kompetenzen, die Lernenden mit wahrscheinlich generell besseren kognitiven Voraussetzungen deutlich besser bei diesem Test abschnitten. Dies gilt jedoch nicht für die starken Lernenden mit viel CLIL - diese erzielten bereits sehr gute Resultate bei der ersten Durchführung und konnten sich deshalb oder auch aufgrund der bereits angesprochenen tieferen Effort bei der zweiten Durchführung kaum steigern. Dass CLIL vielleicht doch einen positiven Einfluss auf das Problemlöseverhalten hat, könnte mit den hohen Mittelwerten der durchschnittlichen und lernstarken Kinder mit CLIL erklärt werden, die nur vergleichbar mit den lernstarken ohne CLIL sind; oder auch mit der Tatsache, dass die Lernenden mit CLIL zu Beginn signifikant besser abschnitten (vgl. Abbildung 12).

Da bei der Gruppe mit hohem CLIL-Anteil auch zwei 5. Klassen dazugehören, sollten die Ergebnisse auch noch innerhalb einer Klassenstufe analysiert werden. Das wurde mit den nachfolgenden zwei Fragen adressiert.

- 4) Wie haben sich die Ergebnisse bei den vergleichbaren Gruppen an 6. Klasskindern innerhalb eines Schuljahres verändert?

Betrachtet man nur die 6. Klasskinder – welche sich am ehesten in zwei vergleichbare Untergruppen bestehend aus jenen mit CLIL und jenen ohne CLIL eingeteilt werden können – zeigt sich erneut, dass bei der CLIL-Gruppe die lernschwachen Schüler:innen beim «Lesen» und «Hören» signifikante Fortschritte erzielen konnten. Da sich jedoch auch die lernschwachen und durchschnittlichen Schüler:innen ohne CLIL im Bereich «Hören» signifikant steigern konnten, ist das noch nicht Beweis genug, dass zusätzliche CLIL-Lektionen diese Kompetenzbereiche per se stärkt. Da die Schreibaufgabe Ende Schuljahr anscheinend anspruchsvoller war als noch zu Beginn des Schuljahres, konnten auch bei dieser Betrachtungsweise nur die lernstärksten 6. Klasskinder mit CLIL bessere Resultate erzielen. Beim Problemlöseverhalten steigerten sich die mid-attainers beider Untergruppen sowie die high-attainers ohne CLIL signifikant. Die durchschnittlichen und lernstarken CLIL-Lernenden zeigen jedoch insgesamt die deutlich besten Ergebnisse bei diesem Test – was vielleicht wirklich darauf zurückzuführen ist, dass CLIL ihnen in diesem Bereich Vorteile verschafft. Schliesslich müssen sie im fremdsprachlich geführten CLIL-Unterricht konstant mit dem «Problem» zurechtkommen, Inhalte zu verstehen und trotz limitierten Englischkenntnissen ihr Lernen und Denken mitzuteilen.

- 5) Wie haben sich die Ergebnisse der 6. Klasskinder mit CLIL im Englisch insgesamt und beim «Tower of London» im Vergleich zu jenen ohne CLIL innerhalb eines Schuljahres verändert?

Sowohl CLIL- als auch non-CLIL-Schüler:innen der 6. Klasse konnten ihre Englisch- und Problemlösungsfähigkeiten innerhalb eines Schuljahres verbessern. Im Fach Englisch zeigten die CLIL-Schüler:innen signifikante Fortschritte mit mittlerer Wirkung. Non-CLIL-Schüler:innen hingegen machten keine signifikanten Fortschritte. Diese Ergebnisse deuten darauf hin, dass CLIL bereits innerhalb eines Schuljahres einen bedeutenden Vorteil für die Entwicklung der Zielsprache bietet. Diese Ergebnisse werden durch frühere Studien bestätigt, die zeigen, dass ein intensiverer Kontakt mit der englischen Sprache tatsächlich einen positiven Einfluss auf die Sprachleistungen der Lernenden hat (z.B. Bauer-Marschallinger et al., 2023; Ohlberger & Wegner, 2018). Da der Effekt jedoch nur mittelstark ist, lässt sich argumentieren, dass ein Schuljahr möglicherweise nicht ausreicht, um offensichtlichere Vorteile zu beobachten (Pérez-Vidal & Roquet, 2015). Im Gegensatz dazu zeigten beide Gruppen eine signifikante Verbesserung ihrer Problemlösungsfähigkeiten, gemessen anhand der «Tower of London» Tests, wobei der Effekt bei den non-CLIL-Schüler:innen stärker war als bei den CLIL-Schülern. Auch bei dieser Betrachtungsweise konnten somit der erhoffte positive Einfluss von CLIL auf die Problemlösungsfähigkeiten nicht eindeutig beobachtet werden. Die Tatsache, dass sich die Problemlösungsfähigkeiten in der non-CLIL Gruppe vergleichbar oder sogar stärker entwickelten, stützt die Annahme, dass diese Fähigkeiten wahrscheinlich eher durch die allgemeine kognitive Entwicklung beeinflusst wurden, die Kinder in diesem Alter innerhalb eines Schuljahres durchlaufen, als durch die zusätzlichen CLIL-Stunden. Dass alle Lernenden am Ende des Schuljahres bei diesem Test deutlich bessere Leistungen erzielten, lässt sich auch dadurch erklären, dass alle Schüler:innen bei der zweiten Erhebung mit der Aufgabe vertraut waren. Das bedeutet, dass die Schüler weniger kognitive Kapazität für das Verstehen der Anweisungen oder der Aufgabe benötigten, sondern sich auf die Lösung des eigentlichen Problems konzentrieren konnten, was folglich ihre Punktzahl verbesserte. Interessanterweise erzielten die CLIL-Schüler:innen jedoch zu Beginn des Schuljahres deutlich höhere Ergebnisse. Daher gab es zu Beginn offensichtliche Unterschiede zwischen den Gruppen, die vielleicht doch auf ihre vorherige Erfahrung mit CLIL zurückzuführen sein könnten. Um solche Vermutungen zu bestätigen, wären jedoch weitere Untersuchungen erforderlich.

Limitationen

Unglücklich scheint mir, dass diese zweite Erhebung ganz am Ende des Schuljahres, kurz vor Klassenlager und mit wenig Energielevel noch diese Testreihe durchgeführt wurde. Insbesondere bei den 6. Klasskindern mit CLIL zeigte sich eine gewisse Schulumüdigkeit. Bei den 6. Klasskindern ohne CLIL wurden die Tests zwei bis drei Wochen früher angesetzt, da waren sie noch mehr im Schularbeitsmodus. Auch ist davon auszugehen, dass die Auswahl der Testaufgaben die Ergebnisse stark beeinflusst: Einerseits waren die standardisierten Tests im Bereich «Sprechen» und «Schreiben» am zweiten Erhebungszeitpunkt angeblich schwieriger; andernfalls hing es bei allen Tests stark davon ab, ob die Lernenden mit den alltagsbezogenen Themen vertraut sind und dazu benötigten Wortschatz kannten. Ebenfalls ungünstig ist, dass das grosse Sampling von fast 200 Kindern aus unterschiedlichen Klassenstufen bestand. Auch wenn mit der Aufschlüsselung der Anzahl Wochenlektionen die Kontaktzeit mit Englisch unabhängig von der Klassenstufe quasi neutral betrachtet werden konnte, so schien die kognitive und generelle Reife, sowie ein Jahr mehr «formaler Englischunterricht» oder auch allgemein mehr Schulerfahrung doch auch einen starken Einfluss auf die Ergebnisse zu haben. Deshalb wurde die weitere

Differenzierung der Schüler:innen nach 5. und 6. Klasse vorgenommen, respektive die Ergebnisse nur für alle 6. Klasskinder als weitere Untergruppe weiterausgewertet. Dies hatte jedoch den grossen Nachteil, dass die Anzahl von Lernenden pro Untergruppe sank und somit die statistische Power beeinträchtigt wurde. So sind jene Aussagen zu den 6. Klasskindern zwar vergleichbar, jedoch nicht mehr so aussagekräftig. Mit der letzten, fünften Frage konnte die Aussagekraft wieder erhöht werden, indem die CLIL und non-CLIL Gruppe insgesamt ohne weitere Aufschlüsselung der Levels betrachtet wurde – was dann schliesslich zwar zu aussagekräftigen, doch weniger ausdifferenzierten Resultaten führte.

Schlussbemerkungen

Als CLIL-Expertin bin ich überzeugt, dass diese Unterrichtsform einen positiven Einfluss auf das Lernen der Kinder hat – angefangen bei motivationalen Aspekten bis hin zum fremdsprachlichen Lernen und auch in Bezug auf allgemeine fürs 21. Jahrhundert relevante Kompetenzen wie Mehrsprachigkeit, interkulturelles Bewusstsein und Problemlöseverhalten. Dass all diese Vorteile zwar auf deskriptiver Ebene zum Vorschein kamen, sich jedoch nicht nur eindeutig bei dieser Untersuchung nachweisen liessen, ist natürlich als Forscherin etwas enttäuschend. Der gesamte Bericht mit all seinen Teilergebnissen und unterschiedlichen Betrachtungsweisen zeigt eindrücklich auf, wie komplex Unterricht ist und welche weiteren Faktoren Lernen – bewusst oder unbewusst - beeinflusst. Kurz: Forschung im Schulfeld findet nie unter Laborbedingungen statt und ist etwas «messy» - das führt die vorliegende Erhebung deutlich vor Augen.

Trotzdem stimmen mich drei Ergebnisse besonders optimistisch: Erstens, dass das innovative Schulmodell mit dem erweiterten Englischunterricht bei den Schüler:innen in Buchrain sehr gut ankommt und sie deshalb auch sehr offen und motiviert der englischen Sprache begegnen. Zweitens, dass einzig die 6. Klasskinder mit CLIL im Vergleich zu jenen ohne CLIL bei den Englishtests insgesamt (vgl. Frage 5) signifikante Fortschritte erzielen konnten, weist daraufhin, dass die zusätzliche Kontaktzeit mit Englisch in einem handlungs- und anwendungsorientierten Setting dem Fremdsprachenlernen einen Boost gibt. Drittens, dass oft die Lernenden mit eher nachteiligen Leistungsvoraussetzungen (z.B. low-attainers) im Englisch die grössten Fortschritte machen konnten, lässt vermuten, dass ein Lernsetting, bei dem die Kommunikation und weniger die Korrektheit im Vordergrund steht, gerade den anderweitig benachteiligten Englischlernenden entgegenkommt. Ohne Zweifel glaube ich deshalb, dass CLIL der richtige Weg ist, das schulische Fremdsprachenlernen weiter zu stärken.

Ich danke der Schule Buchrain und allen beteiligten Personen für die angenehme Zusammenarbeit.

Silvia Frank Schmid

Luzern, 26. August 2025

Literaturhinweise

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Second Edition). Lawrence Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Kaller, C. P., Unterrainer, J. M., Kaiser, S., Weisbrod, M., & Aschenbrenner, S. (2011). *Manual: Tower of London—Freiburger Version*. Schuhfried GmbH.
- Ohlberger, S., & Wegner, C. (2018). Bilingualer Sachfachunterricht in Deutschland und Europa. *Herausforderung Lehrer*innenbildung - Zeitschrift zur Konzeption, Gestaltung und Diskussion*, 45-89 Seiten. <https://doi.org/10.4119/HLZ-2390>
- Pérez-Vidal, C., & Roquet, H. (2015). CLIL in Context: Profiling Language Abilities. In M. Juan-Garau & J. Salazar-Noguera (Hrsg.), *Content-based Language Learning in Multilingual Educational Environments* (Bd. 23, S. 237–255). Springer International Publishing. https://doi.org/10.1007/978-3-319-11496-5_14
- Schwarz, J., Heidi Bruderer Enzler, Schwarz, J., & Käch, W. (2023). *Universität Zürich: Methodenberatung*. https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/zusammenhaenge.html#11