DE GRUYTER MOUTON

Emmanuelle Le Pichon*, Sergio Baauw, Sohee Kang and Jacob Vorstman

School Trajectory of Elementary Newcomer Students: Early Tracking System and Norm Accommodation

Abstract: The increase in the number of newcomer students in countries across the world has underscored the importance of effective transition strategies in education. Many students encounter difficulties in acquiring academic knowledge due to initial limited language skills in the school language. Implementing appropriate strategies to facilitate this transition has shown positive outcomes. However, some obstacles remain to be addressed for newcomer students. In the Netherlands, standardized tests are utilised to monitor the academic progress of all students and determine the appropriate educational pathway. Our study, based on longitudinal assessment data from 51 newcomer elementary school students and 74 of their classmates, sheds light on how this system interacts with the transitions of newcomers. While newcomer students made greater progress compared to their peers, we found that their results were influenced by the application of test norms designed for younger age groups. The insights from this study provide valuable perspectives on educational pathways for newcomer students and prompt us to reconsider the implications of norm accommodation for these students. It highlights the needs to implement practices that enable newcomer students to effectively pursue their academic aspirations.

Keywords: assessment, norm accommodation, multilingual learner, newcomer student, academic performance

*Corresponding author: Dr. Emmanuelle Le Pichon, University of Toronto, Ontario Institute for Studies in Education, Toronto, Canada, E-Mail: e.lepichon@utoronto.ca
Dr. Sergio Baauw, Utrecht University, Utrecht Institute of Linguistics, Utrecht, Netherlands, E-Mail: s.baauw@uu.nl
Dr. Sohee Kang, University of Toronto Scarborough, Computer and Mathematical Sciences, Toronto, Canada, E-Mail: sohee.kang@utoronto.ca
Jacob Vorstman, Hospital for Sick Children Toronto, Department of Psychiatry, Toronto, Canada, E-Mail: jacob.vorstman@sickkids.ca

Résumé : L'augmentation du nombre des élèves primo-arrivants dans les pays du monde entier a révélé l'importance de la mise en place de stratégies de transition pédagogiques appropriées. Nombre de ces élèves rencontrent des difficultés dans leur apprentissage scolaire, en raison de leur connaissance initialement limitée de la langue de l'école. La mise en œuvre de stratégies appropriées pour faciliter cette transition a donné des résultats positifs. Toutefois, certains obstacles restent à surmonter. Aux Pays-Bas, des tests standardisés sont utilisés pour suivre les progrès scolaires de tous les élèves et déterminer le parcours éducatif approprié. Notre étude, basée sur les données d'évaluation longitudinale de 51 élèves primo-arrivants et de 74 de leurs camarades de classe, met en lumière la manière dont ce système interagit avec les transitions des primo-arrivants. Alors que les élèves primo-arrivants ont progressé davantage que leurs pairs, l'application de normes de test conçues pour des groupes d'âge plus jeunes a considérablement influencé leurs résultats. Les conclusions de cette étude apportent des éclairages importants sur les parcours éducatifs des élèves primo-arrivants et nous incitent à reconsidérer les implications de l'adaptation des normes pour ces élèves. Elle met en évidence la nécessité de mettre en œuvre des pratiques qui permettent aux étudiants primo-arrivants de poursuivre efficacement leurs aspirations académiques.

Samenvatting: De toename van het aantal nieuwkomers in landen over de hele wereld heeft het belang van effectieve overgangsstrategieën in het onderwijs onderstreept. Veel studenten ondervinden in het begin problemen bij het verwerven van academische kennis als gevolg van hun aanvankelijke beperkte taalvaardigheid in de schooltaal. Het implementeren van geschikte strategieën om deze overgang te vergemakkelijken heeft positieve resultaten laten zien. Sommige obstakels moeten echter nog worden aangepakt voor nieuwkomerstudenten. In Nederland worden gestandaardiseerde toetsen gebruikt om de academische vooruitgang van alle leerlingen te volgen en het juiste onderwijstraject te bepalen. Onze studie, gebaseerd op longitudinale toetsgegevens van 51 nieuwkomerleerlingen op de basisschool en 74 van hun klasgenoten, werpt licht op hoe dit systeem interageert met de transities van nieuwkomers. Hoewel nieuwkomersleerlingen meer vooruitgang boekten in vergelijking met hun leeftijdsgenoten, ontdekten we dat hun resultaten werden beïnvloed door de toepassing van testnormen die zijn ontworpen voor jongere leeftijdsgroepen. De inzichten uit dit onderzoek bieden waardevolle perspectieven op onderwijstrajecten voor nieuwkomers en nopen ons de implicaties van normaanpassing voor deze leerlingen te heroverwegen. Het benadrukt de noodzaak om praktijken te implementeren die nieuwkomerstudenten in staat stellen om hun academische aspiraties effectief na te streven.

1 Introduction

1.1 Increased migration poses challenges to the school system

The profound societal change caused by the influx of newcomer students throughout Europe in recent years has created an urgent need to identify the most successful strategies geared towards a successful transition of these students to the educational system (Koehler et al., 2018; Bravo-Moreno, 2009). The term "newcomer students" is used in education policy and academic discourse to designate students arriving from a different country at school age and without, or with insufficient knowledge of the school language (Kieffer & Thompson, 2018). The transition these students are required to make is highly challenging and consists of adjusting to a new educational, linguistic and cultural environment while at the same time trying to acquire academic content as much as possible (Griebel et al., 2013). Not surprisingly therefore, as it was pointed out more than 20 years ago, this transition, if not addressed adequately, may potentially cause problems related to social and academic functioning (Galton, Gray & Ruddock, 1999). The issue is highly relevant as studies have shown that newcomer students have a higher risk of truancy and dropping out of school early without a diploma compared to the rest of the student population (Anisef et al. 2010; Traag, Rolf & van der Velden, 2011; Koehler, Palaiologou & Brussino, 2022). *Vice versa*, the use of adequate strategies to manage this transition is associated with more positive outcomes (Estrada & al. 2016; Kieffer & Thompson, 2018). For example, adequate schooling during the early years of resettlement improves long-term adjustment, the school playing an essential role in socialisation and integration of newcomer students in society (De Wal Pastoor, 2015). Important questions, then, are what factors contribute to an optimal transition during the early years of resettlement, and how to translate these insights into policies that can influence the schooling systems (Koehler & Schneider, 2019; Estrada et al. 2016; Bravo-Moreno, 2009). Today, numerous studies have contributed to a better understanding of the impact of educational language policies to facilitate the transition of newcomer students. However, the variability across support strategies for newcomer students, even within the same country, complicates the identification of these factors. In addition, newcomer students represent a vastly heterogeneous group (Browder, 2018; Kieffer & Thompson, 2018). Although all students in this group share with each other that they have arrived from a different country in a new school system, they differ from each other in terms of age, country of provenance, previous schooling history, language repertoire, and personal experiences of migration, which renders comparisons across studies extremely difficult (Herzog-Punzenberger, Le Pichon-Vorstman & Siarova, 2017).

1.2 Specific reception classroom programs for newcomer students

In most instances, newcomer students are required to learn the school language as soon as possible in order to continue their acquisition of academic content of the major school subjects (Le Pichon, 2020a&b, Le Pichon et al. 2020). For this reason, most Western European countries have introduced a preparatory class, purportedly to allow for an emphasis on expedited learning of the language of instruction before entering mainstream classrooms of the regular school system (Le Pichon, 2020 a & b, Le Pichon & Baauw, 2020; Koehler & Schneider, 2019; Koehler, Palaiologou & Brussino, 2022). These language preparatory classes may be part-time or full-time, and may be organised in separate schools, or in classrooms of regular, mainstream schools (Dryden-Peterson et al. 2019). Additionally, some countries differentiate between students with and without schooling experience prior to their arrival in the country in question, this, for example, is the case for Ontario, Canada (Gagné, Schmidt & Markus, 2017). Others, such as the Netherlands, only consider mastery of the language of schooling as a relevant factor for the orientation towards these preparatory classes. In the Dutch system, most students who arrive from a country whose language is not the language of the school will be oriented for at least one year to this preparatory system.

In addition, the duration of the preparatory classes varies considerably; while the general concept of the language preparatory strategy is shared between many countries, there is no consensus on the optimal duration of these classrooms and even less on the curriculum or on the pedagogy to adopt, both of which, consequently, varies greatly between and even within countries. The Dutch system outlined above is not implemented everywhere in the country; depending on the region in which newcomers are educated, they may be integrated into regular classes without special support, sent to a preparatory class within a regular school for one to two years, or sent to a preparatory class in a separate school for two years, as is the case in the study reported here (Koehler & Schneider, 2019; Le Pichon, 2020; Koehler, Palaiologou & Brussino, 2022). In Belgium, like in France, pulling students out of the classroom for a few hours per day is the preferred strategy, allowing students to join regular classes for courses that are deemed less demanding in terms of language (e.g., physical education). However, in the Flemish part of Belgium, local schools have a large degree of autonomy with regard to the organisation of newcomer education. In primary education, schools can either include the non-Dutch speaking newcomers in an existing mainstream class, or place them in a separate newcomer class, or combine both options. In secondary education, newcomer students are placed in a reception class for a year, in which the focus is on learning Dutch and integration into Belgian society. In sum, despite a broad consensus on the need to prepare newcomer students, there is still much divergence

regarding which strategies are implemented (Coelho, Oller & Serra, 2011; Thompson, Umansky & Porter, 2020).

1.3 Educational achievement of newcomer students

Despite these complexities inherent to the population under study and the scarcity of academic research on the educational achievement of newcomer students (Koehler, Palaiologou & Brussino, 2022), some findings start to converge across studies: Upon their arrival in the new school, newcomer students usually lag behind their peers (Browder, 2018; Kieffer & Thompson, 2018). The performance gap is more common for those who speak a different language at home (other than the language of instruction) and for those in a disadvantaged socio-economic situation (OECD, 2010: 37–38). This gap is even more pronounced in countries with an early tracking system like the Netherlands (Werfhorst, 2018). However, these findings are not unequivocal as some countries report higher or equal school achievements of this particular population than the average population. This is the case in studies from Australia (Balogh, 2016), Canada (Volante et al. 2017) or recently, from the United States (Kieffer & Thompson, 2018). Hypothesised explanations for those contradictory findings include differences in socio-economic status, country of provenance, schooling system of the country of arrival as well as methodological differences between various studies.

In this article, we aim to contribute to the understanding of strategies to support the transition of newcomer students. To this end, we set out to examine academic performance in a group of newcomer students. To address at least one source of variability, we conducted this study in one country (the Netherlands) and recruited newcomer students after completion of the same preparatory school. We collected longitudinal standardised academic test results taken at their start in a mainstream Dutch school ("baseline"; i.e. immediately after completing the language school trajectory), and after an average follow up time of 21 months ("follow-up"). We used as a comparison the same scores from a group of classroom peers of the same newcomer students. The analysis of the data sets obtained from schools was done on the basis of anonymized data for the newcomer students and their peers and informed consent was obtained at the time of initial data collection. While analysing the results we noticed that in many instances, the standardised academic test results reported by teachers were obtained by applying test norms appropriate for younger age groups, a practice called "norm accommodation". Subsequent to this observation, we formulated a new (additional) objective to the study, namely, to examine extent and impact of norm accommodation on the measures of academic performance in newcomer students, and in their classroom peers.

In this manuscript, we start by explaining the concept of norm accommodation and outline our different analytical strategies to quantify both its application and its impact on test scores. The results we report are pertinent to our original objective, i.e. to examine the trajectories of academic performance in newcomer students and their classroom peers. In addition, and in line with the added objective, we will highlight how these observations are influenced by the phenomenon of norm accommodation.

1.3.1 Norm accommodation

Norm accommodation is relevant in those situations where test results, obtained through some form of standardised assessment of academic performance, require contextualisation using age-specific norm data. In this context, norm accommodation can be understood as the practice of applying age-specific norms that deviate from the actual chronological age of the student. The phenomenon of norm accommodation can be considered and quantified from two perspectives:

–   First, *the extent* with which it is applied, quantified as the difference between the actual age of the student and the target age of the norms that were used to quantify test performance.
–   Second, *the impact* of norm accommodation on the resulting score, referring to the degree to which the resulting performance scores are altered by the applied accommodation.

2 Hypotheses

This study was set out to examine the academic performance of newcomer students in elementary school after completion of their reception classroom year and integration into a mainstream classroom environment. In line with the initial objective of the study, as well as the additional objective concerning norm accommodation, we formulated the following hypotheses:

1.   At baseline, the newcomer students perform academically below the level of their classroom peers;
2.   Longitudinal trajectory of results will reveal a catch-up of academic performance in the newcomer group as they improve their versatility in the language of the school;
3.   Norm- accommodation is applied broadly but mostly in the newcomer group;
4.   Under the assumption that norm accommodation is an adequate strategy to support students who are still learning the language, we expect that the necessity of norm accommodation will decrease as a function of time. Consequently, we hypothesize that longitudinal results will show a decrease in norm accommodation.

Given that results pertinent to the first hypotheses 1 and 2 are likely influenced by the impact of norm accommodation (hypothesis 3), we start by reporting the results pertaining to norm accommodation, and subsequently report our observations with regard to the first two and the last hypotheses.

3 Research methods

3.1 Research setting: newcomer education in the Netherlands

At the primary school level, Dutch municipalities organise the schooling of newcomer students according to different models. The newcomer students included in the current study had all been schooled in the same separate language school specialised in the reception of elementary newcomer students. After approximately one year of language preparatory schooling, newcomer students were transferred to a mainstream elementary school, usually in their own neighbourhood.

3.2 Participants

Classroom peers were selected randomly from the mainstream elementary classrooms in which the newcomer students were placed, using a random number generator. The total sample included 125 participants of whom 51 were newcomer students and 74 classroom peers (see table 1). Of these, in 117 participants data was available for both time points for our longitudinal analyses. Since results were not available for all students for all five tests, there is some variation in the exact number of results per test.

Table 1: demographics of both groups and time points 1 and 2

		All participants	Newcomer students	Classroom peers
	n	125	51	74
Baseline (age in months)	Mean age (months) ± stdev	105.7 ± 18.8	105.5 ± 17.6	105.9 ± 19.7
	Age range (months) (min – max)	75.0–152.0	76.0–141.0	75.0–152.0
Follow-up (age in months)	Mean age (months) ± stdev	126.2 ± 18.0	127.1 ± 18.2	125.6 ± 17.8
	Age range (months) (min – max)	81.0–162.0	81.0–159.0	86.0–162.0

3.3 Measure of academic performance

Most primary schools in the Netherlands use the Central Institute for Test Development (CITO) tests to monitor the progress of students. The assessment starts in grade 1 of the Dutch school system, and continues until grade 6, the last group of primary education. There are specific tests for the different school skills: Mathematics, Text Reading (so-called AVI test), Three-minutes-words Reading, Technical Reading, Reading Comprehension, Spelling and Vocabulary. Per school year, 2 tests are taken per skill, the M(idden)-test, which is taken in the middle of the school year, and the E(ind)-test, which is taken at the end of the school year. Raw test scores are converted to scaled scores which allows comparison with age-specific norms available for each test moment across the elementary school trajectory (e.g., tests performed in the middle of grade 1 or at the end of grade 5, can be compared after conversion to the norms provided for "M3" or "E5" respectively) (HCO, 2016)[1]. Finally, the result can be converted into five percentile categories: Score A refers to the top 25 % highest scores of the student population, B for the 25 % who score well to just above the national average, C stands for the 25 % that scores just to well below the average of the national population, D represents the 15 % that scores well below the national average, and E represents the bottom decile of 10 % lowest scoring students. Taking a conservative approach, we considered scores D and E to as indicators of "insufficient" grades, while we considered scores A, B and C as in the "sufficient" range.

To investigate the development of the newcomer students, we analysed the results of assessments acquired at two time points; first when the newcomer students joined the mainstream school (baseline), and second at a fixed time point, (on average) 21 months after their start in the mainstream school (follow-up). Longitudinal data from two-time points allowed us to measure development. The peers were paired to the newcomer students with respect to the two test moments so that the time interval between first and last test moment of the peers was equal to that of the newcomer students. As can be seen from table 1, the age at the first and second test moment were relatively comparable between both groups. Thus, in this study, we included data from two different time points: the end time point (follow-up) can be considered cross-sectional because the results of the second series of assessments were all derived at a fixed time point, namely the winter of 2014; the baseline time point is retrospective and varies as a function of the first moment the newcomer students were assessed in their new school.

---

1 https://dokumen.tips/documents/tabellen-tussenopbrengsten-cito-lovs-groepsniveau.html

3.3.1 Five modalities

Results of these tests included five modalities: 1. Mathematics, 2. Reading comprehension, 3. Three-minutes-words reading, 4. Spelling and 5. Vocabulary.
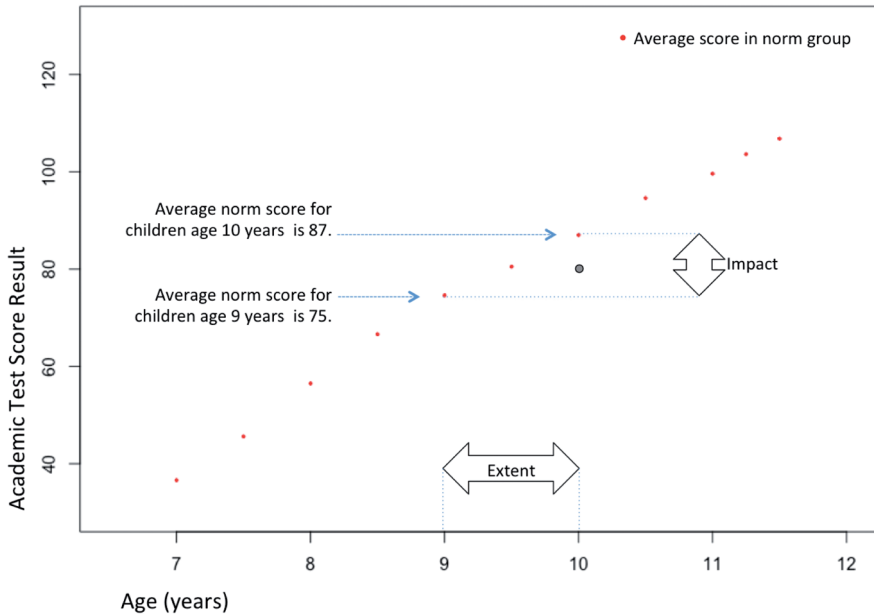

Figure 1: Imagine a 10-year-old who achieves 80 (grey circle) on an academic performance test. A standardised score from this result can be generated by placing this result in a distribution of results obtained in a norm group of students of 10 years. In this example, the student scores 7 points lower than the average of other students of the same age (the norm average for this group). Norm accommodation occurs when the student's score is standardised in a distribution of scaled scores obtained in a group of students of a different age. For example, when a teacher decides to use the norm group for age 9, the average of which is 75. The *extent* with which the teacher applies norm accommodation can be expressed as the age difference between actual age (10) and the age of the applied norm (9), which in this example would be minus 12 months. The *impact* of norm accommodation can be expressed as a 12-points difference given that with norm accommodation, the student's academic performance can be quantified as 5 points above average, while without norm accommodation the student this measure would be 7 points below average.

The resulting scaled score for each test can be expressed quantitatively as a function of its distance to the average norm score for that group, as illustrated in figure 1: in this example the student achieved a scaled score of 80 on the test; this can be expressed as (a distance of) "-7" relative to the average score in the norm group for 10-year-olds. Alternatively, the same scaled score can be expressed as "+5" relative to

the norm for students aged 9 years as the score is 5 points above the average norm score in this age group. See below for more explanation on the concept of norm accommodation.

3.3.2 Data reduction

At baseline and follow-up time points of the study, we obtained for each study participant academic performance scores for these five test modalities. In order to avoid multiple testing, we performed a principal component analysis (PCA)(Johnson & Wichern, 2007), using as input the scaled scores relative to the norm average. The PCA allowed us to use one single measure of academic performance (the first principal component, PC1) rather than five measurements per individual. Detailed results of the PCA are provided in our supplemental material. In short, the first two principal components together explained 7 % of the variability of multivariate data as reported by the schools. We repeated the PCA after applying age norms consistent with the students' actual age (i.e., without norm accommodation) and found highly similar PCA results, with 81 % of the variability of the data explained by the same first two PCs. In both PCAs, the first PC explained most of the variability of the data (55 % and 68 %, respectively) with same-direction effects by all five tests, but most prominently math, reading comprehension and vocabulary (see supplemental material). Given these investigations, we considered the first PC as a robust measure of academic performance for our downstream analyses.

3.4 Norm accommodation: application and impact

The method of quantifying the *extent* with which norm accommodation is applied is straightforward and consists of subtracting the actual age of the student from the target age of the applied norm average. When norms are provided for grade levels instead of age, as is the case for the Dutch CITO tests (CITO, 2019), the target age can be inferred from the average age that is typical for a certain grade. For ease of reading, the authors have converted the Dutch levels into levels corresponding to the most common system (kindergarten to 12).

The *impact* of norm accommodation can be quantified by comparing the academic performance measure based on the accommodated norm average versus that based on the norm average for the actual age of the student at the time of assessment. Given that math, reading comprehension and vocabulary are the main drivers of academic performance (the first PCA cluster; see supplemental material), we based our calculations of norm accommodation effects based on results of these three tests.

3.5 Academic performance: longitudinal trajectory

Hypotheses 2 and 3 both allude to the longitudinal academic trajectory. To this end, our analyses focused on a comparison of academic performance at the first time point (baseline, abbreviated B) and endpoint (follow-up, abbreviated E). We accounted for variability in the duration of the interval trajectories between baseline and follow-up by dividing delta (B-E) academic performance by the number of months of the time interval (B-E) for each individual. For example, a performance measured as -10 at baseline and +14 at follow up indicates an advance of 24 points. If one student accomplished this in 8 months (i.e., time between B and E was 8 months), the academic progress would be quantified as 3 (24 points divided by 8 months). If a different student made the same progress in 16 months, the progress would be expressed as 1.5 (24/16).

3.6 Academic performance: dimensional and binary perspective

For our study on application and impact of norm accommodation (hypotheses 3 and 4), we examined academic performance, the score generated by the first PC in a dimension reduction way (henceforward referred to as "academic performance"). For hypothesis 3, we examined the mean of the academic performances obtained at both time points (B and E). We compared the distribution of academic performance scores with and without norm accommodation using the Welch two Sample t-test. To compare the effect of norm-accommodation between newcomer students and their peers, we calculated the difference between the group average of academic performance obtained with and without norm accommodation in both groups. Given a non-normal distribution of this variable, we applied the Wilcoxon rank sum test to test whether any observed difference was statistically significant.

To test our last hypothesis (4), we investigated a putative change of norm accommodation over time. Therefore, we examined the difference, rather than the mean, between the academic performance at baseline and follow-up for this analysis, corrected for time interval variability as explained previously.

In addition to the PCA generated scores, we also examined the results from a categorical perspective. Here, our motivation was to ensure that our findings can be interpreted such that it has direct relevance to the actual school context. To this end, we distinguished between those who achieved a grade in the sufficient to high range (i.e., C, B or A) and those who scored in the insufficient grade range (i.e., D or E). Subsequently, we compared the proportion of students with sufficient grades for the three main test modalities driving academic performance (math, reading com-

prehension and vocabulary) between the newcomer student and the classroom peers, and the change in this distribution at baseline and follow-up. We provide these results as explorative, i.e., to generate hypotheses for future studies and therefore refrain from formal statistical testing.

4 Results (part I, pertinent to hypotheses 3 and 4)

Results relating to hypothesis 3 related to the impact of norm- accommodation (see fig 2) show that, first, in both the newcomer and the peer group, the effect of norm accommodation on academic performance was substantial, with significantly higher performance results with norm accommodation compared to without norm accommodation (Welch Two Sample t-test; $p=1.87*10^{-5}$ and $p = 0.01$, respectively).

Second, while in both groups, the impact of norm accommodation on academic performance was significant, the effect size was higher in the newcomer group compared to the peer group (Wilcoxon rank sum test; $p = 0.028$).
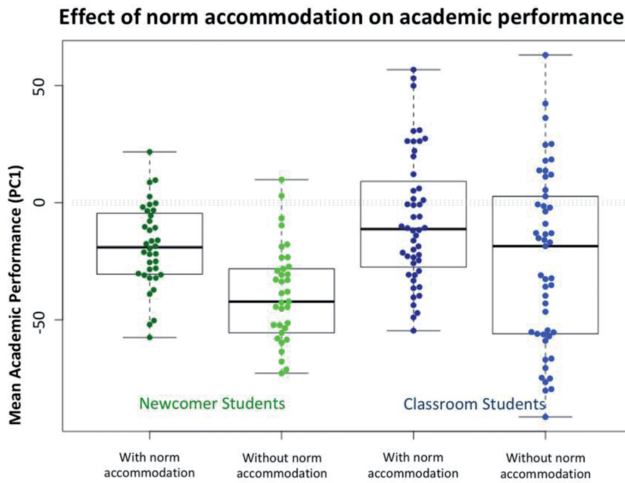

Figure 2: Academic performance with and without norm accommodation. Each dot represents one participant. The Y axis displays the scaled score relative to the norm average, with 0 indicating that the student scored equal to the norm average (dotted horizontal line). Academic performance scores achieved by newcomer students are in green and those by classroom peers are in blue. For each group the academic performance measures with (dark) and without (light) norm accommodation is represented by different shades of green and blue respectively. In both groups, the impact of norm accommodation is significant ($p=1.87*10^{-5}$ and $p = 0.01$, respectively), but the impact is larger in the newcomer group compared to peers as is evident by the larger difference of the distributions with and without norm accommodations in the newcomer group ($p = 0.03$).

In addition to examining and comparing *the impact* of norm accommodation, we also analysed *the extent* to which teachers accommodated the norms in both groups. We expected that age accommodation would decrease as a function of the time the student is in school.
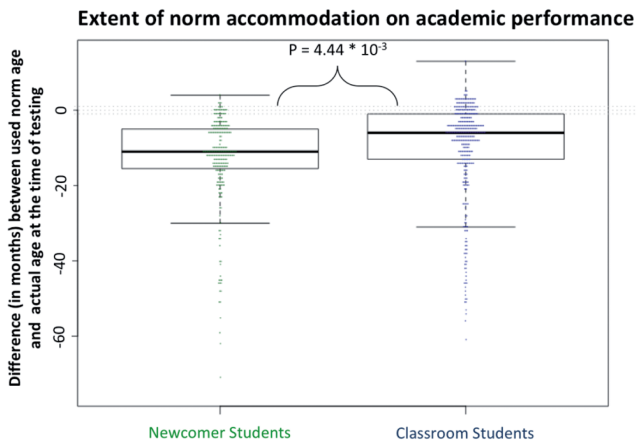

Figure 3: The extent to which norm accommodation was applied, expressed as the difference between the used norm age and the actual age of the student at the time of testing (in months). Here, the horizontal dotted line indicates that no accommodation was used (i.e., the difference is 0 months). Each dot represents norm accommodation applied for the three tests driving academic performance in the PCA (math, reading comprehension and vocabulary). Academic performance scores achieved by newcomer group of students are in green and those by classroom peers are in blue. The extent to which norm accommodation is applied is significantly higher in the group of newcomer students compared to their classroom peers (p = 0.0044). To avoid undue inflation, students older than 12.5 years were removed prior to the analysis.

First, looking at newcomer and peers together, the average difference between actual age and target age of the applied norm was 12.9 months (range -75 to +13 months).

Second, we found that there were students older than 12 years who were still in grade 6; however, CITO provides no norms beyond the end of grade 6. As a result, the gap between age and target age used for norm standardisation quickly escalates for those students of 12.5 years and older. To avoid undue influence of this ceiling effect, we repeated our analysis after exclusion of students older than 12.5 years in our sample. We found that this intervention altered our initial results only marginally: the average difference between actual age and target age of the applied norm was 11.4 months (range -71 to +13 months).

Third, we compared the extent with which norm accommodation was applied between the newcomer and peer groups. In the newcomer group, the average norm

accommodation applied by teachers was 13.3 months below the actual age of the child (range: -71 months to + 4 months); in the classroom peers this was -10.2 months (range -61 to +13 months). The difference between both groups was significant (see fig. 3, results shown with exclusion of students older than 12.5 years; Welch Two Sample t-test; p = 0.0044).

The wide range of the norm accommodation raises the possibility of data skewing by the extreme outliers. Therefore, we dichotomized test events using conservative cut-off of norm-accommodation of at least 12 months below the actual age of the student at the time of assessment (i.e., close to the observed average norm accommodation of approximately 12 months). We subsequently compared the proportions of test events for which norms were accommodated with at least 12 months (or more) between newcomers and their classroom peers. We observed that this was the case in 140/275 tests performed in the newcomer group (50.1%) compared to 124/382 tests performed in their classroom peers (32.4%) (Pearson's Chi-squared test with Yates' continuity correction; p = $2.90 + 10^{-6}$).

Finally, we examined whether the extent to which norm accommodation is applied changes over time. Our findings showed barely any change in the proportion of tests for which teachers applied a norm accommodation of at least 12 months between the first and the last time point. As reported above, at baseline, the proportion of students is much higher in the newcomer group (~51%) compared to the classroom peer group (~32%). Our results show little to no change in this proportion over time, for both groups (in the newcomer group the proportion stays exactly at 51% at follow-up, in the classroom peers, it increases from 32% to 34%, which is negligible).

In sum, results or our study pertinent to norm accommodation show that:

(1) Norm accommodation has a significant impact on the overall academic performance scores;

(2) Academic performance measures are accommodated using norm averages for age-equivalents of on average a year younger than the actual age at the time of assessment;

(3) Impact on academic performance is observed both in newcomers and classroom peers; however, both <u>extent of norm-accommodation</u> and <u>its effect on performance scores</u> are significantly greater in newcomers compared to peers;

(4) In contrast to our expectations (hypothesis 4), the proportion of tests for which substantial accommodation was applied remained stable (in both groups).

5 Results (part II, pertinent to hypotheses 1 and 2)

Observations regarding the first hypothesis indicate that both with and without norm accommodation, the average performance of newcomer students is below that of their classroom peers (figure 2).

To test the second hypothesis, we examined academic performance progress by multivariate analysis, PCA, calculated as the difference between the first PC (PC1) at baseline and follow-up time points. Using the results as reported by the teachers (i.e., with various degrees of norm accommodation), our findings indicate that the newcomers group made, as expected, on average more progress compared to their classroom peers (Welch Two Sample t-test, $p = 0.011$, see fig. 4). Subsequently, we examined the extent to which this measure of academic progress was affected by norm accommodation. When comparing progress with and without norm accommodation, we observed a negligible, not significant difference in both groups ($p = 0.46$ and $p = 0.91$ in the newcomers and the classroom peers, respectively, see fig. 5).
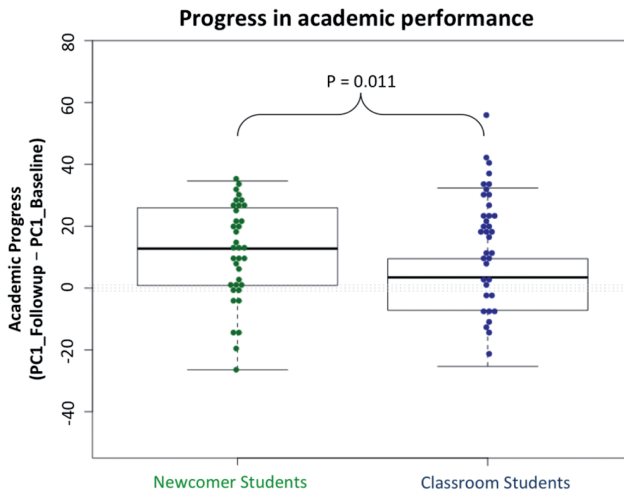

Figure 4: Comparing progress (the difference in score between the first and the last time point; here the horizontal dotted line indicates stagnation (no difference between both time points); positive values indicate advance in academic performance. On average, progress in the newcomer group is significantly higher compared to the classroom peers ($p = 0.011$). For this analysis, we used the results as reported by the teachers.
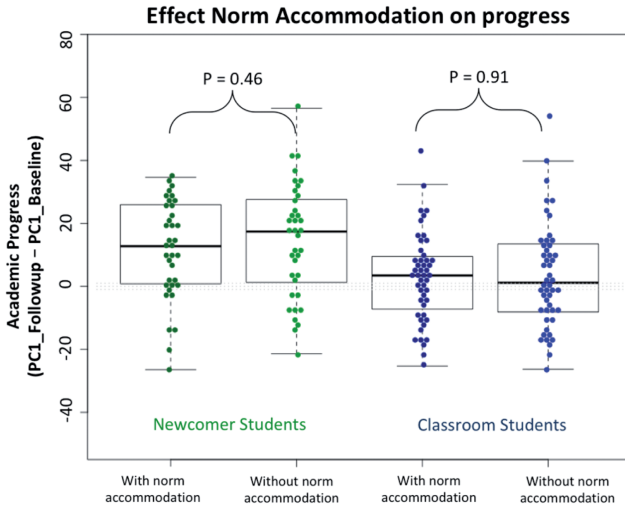
Figure 5: The impact of norm accommodation on progress (the difference in score between the first and the last time point). Here, the horizontal dotted line indicates stagnation (no difference between both time points); positive values indicate advance in academic performance. The figure shows a negligible effect of norm accommodation on progress when expressed as a dimensional outcome (p = 0.46 and p = 0.91 in the newcomer and the classroom peers, respectively)

In summary, the results of our analyses regarding progress of academic performance show that:

(1) Newcomer students perform below the level of their peers on average, but
(2) Newcomer students make more progress than their peers on average and norm-accommodation does not affect this result.

One of the possible consequences of norm accommodation is the development of a more positive view on academic performance than what is actually the case when using age-appropriate norms. Therefore, as the last set of analyses, we examine an outcome that has more real-life relevance, namely the proportion of students scoring grades in the sufficient range (A, B or C). Our observations suggest that norm accommodation had a substantial impact on the proportion of students (newcomers and classroom peers) scoring sufficient grades. Sample size combined with the number of different test modalities restricted our *a priori* power to conduct these analyses in a statistically robust way. Therefore, we did not perform statistical analyses of these results and consider them as exploratory. However, our results highlight the impact of norm accommodation across different test modalities. [See figure 6, 7 and 8].
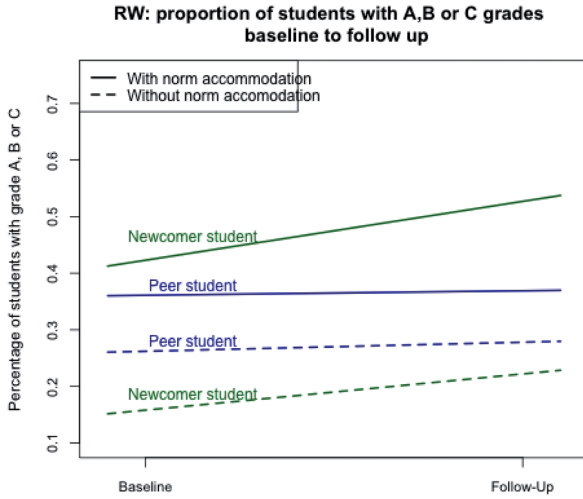
**RW: proportion of students with A,B or C grades
baseline to follow up**

Figure 6: Math

The results as reported by the teachers (i.e., with various degrees of norm-accommodation) suggest that the proportion of newcomer students with sufficient grades is larger than classroom peers at baseline, and that this advantage has increased at follow-up. Without norm accommodation, the situation at baseline is reversed; in both groups the proportion of students with higher grades is lower, but the proportion of students with higher grades is larger in the peer group at baseline. Over time, there is an increase of students with higher grades in the newcomer group, but the proportion remains below that of the classroom peers at the last time point.
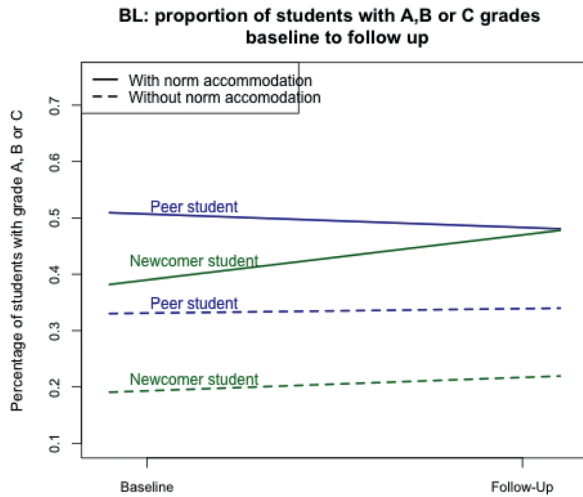
**BL: proportion of students with A,B or C grades
baseline to follow up**

Figure 7: Reading Comprehension

The results as reported by the teachers (i.e., with various degrees of norm-accommodation) suggest that less newcomer students have sufficient grades at baseline, but that they catch up over time such that proportions of higher grades are equal in both groups at the last time point. However, without norm-accommodation, in both groups the proportion of students with sufficient grades is lower and there is no catch up over time, with the difference between both groups essentially unaltered at follow-up.
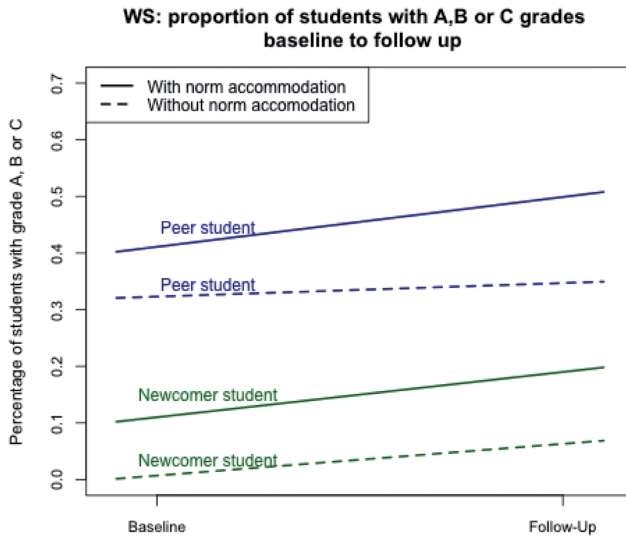

Figure 8: Vocabulary

The results as reported by the teachers (i.e., with various degrees of norm-accommodation) suggest that the proportion of newcomer students with sufficient grades is smaller than that in their peer group, both at the first and at the last measurement, with slight improvement over time in both groups.

Without norm accommodation, the results show a similar pattern, however with overall smaller proportions of students with higher grades in both groups and a slightly blunted progress over time.

6 Conclusion

This study was commissioned by the municipality in which the preparatory school was located to the authors of the present article. The question concerned the effects of this type of preparatory school on the time it takes for newcomer students to

catch up academically with their peers without being constrained by the language to be learned, the groundwork having been laid by Cummins in 1981 and Collier and Thomas in 1989. We therefore retrospectively followed 51 newcomer students who entered the mainstream school system after two years in the same preparatory school. This method allowed us to compare the academic performance of these students with that of their own classmates, i.e., controlling for the type of school in which they were enrolled.

The primary finding is the observation that, in our study, newcomer students make significantly greater progress than their classmates, as expected and consistent with, for example, the study by Kieffer and Thompson (2018). However, a subsequent analysis allowed us to detect the presence of an unexpected practice: teachers often used norm averages that were appropriate for younger students to quantify academic performance of their students. We refer to this phenomenon as "norm accommodation", defined by the application of age-specific norms that deviate from the actual chronological age of the student. We considered this practice in terms of the extent to which it is applied and its impact on the resulting score. We found that newcomer students, after completion of their preparatory school, were evaluated with norm averages below their age. Interestingly, results showed that teachers also used norm accommodation for class peers, albeit to a lesser extent than in the newcomer student group. In addition, contrary to our expectations, we observed that the discrepancy between actual age and the age-equivalent of the applied norms did not diminish over time. Most importantly, the actual extent of the delay of academic performance was mitigated by the effects of norm accommodation.

In the next paragraph, we will discuss the norm accommodation practice. Our purpose is not to denounce this practice but rather to demonstrate that it is frequently used, at least in our data set, and to understand its ramifications for our understanding of the academic progress of newcomer students. We emphasise that these consequences may be relevant for all students, but perhaps particularly for newcomer students who are in the process of catching up academically. Indeed, a number of countries, such as the Netherlands or Germany, still practice early tracking to determine the academic level of the students before entering secondary school. At this point, all students are required to do the same test, the results of which are interpreted centrally; hence a previous "advantage" obtained by norm accommodation is suddenly nullified.

7 Discussion

7.1 What role does accommodation play in supporting language learners?

In a 2004 article, Abedi, Hofstetter and Lord discussed a range of strategies to adapt testing for school language learners as follows: "The goal of [testing] accommodations is to provide a fair opportunity for English language learners to demonstrate what they know and can do, to level the playing field, so to speak, without giving them an advantage over students who do not receive the [testing]accommodation." (Abedi, Hofstetter, & Lord. 2004:2). (Note that Abedi, Hofstetter and Lord consider a broad set of potential adjustments to testing, which is different from the concept of norm accommodation such as reported in the current study). They highlighted the frequency of their widespread use across the United States. However, they went on to ask a number of pertinent and specific questions to understand the situations in which testing accommodations should or should not be applied. It seems important to restate them here: "Does using accommodations yield more valid inferences about an English learner's knowledge? Which students should be eligible and what criteria should be used to decide their eligibility? What type of testing accommodation should be used? Are some testing accommodations more effective than others —and if so, are they more effective in general or only for particular students? Do testing accommodations give students who receive them an unfair advantage? Is it meaningful to compare English learners' accommodated scores with English-proficient students' non-accommodated scores? What implications do test accommodations have for test administration and testing policy more generally?" They also pointed out a lack of research related to these questions but already feared a reduction of quality instruction (2004: 4). Although inclusive intentions are believed to be the reason behind norm accommodation, such as providing positive encouragement to students, it can also pose challenges for their academic progress. Initially, accommodating students is understandable, as second language learners typically require about five years of schooling to perform at their full cognitive level (Cummins, 1981; Collier & Thomas, 1989). Therefore, a temporary setback in academic performance can be attributed to a lack of proficiency in the school language. However, the difference between our research and studies on testing accommodations for newcomer students is that in the latter, testing accommodation is expected to support students by modifying the test to make it more comprehensible from a language point of view allowing for instance for more testing time, bilingual versions of the tests, or modified version of the text. In our study, this is not the case. Norm accommodation in our study relates the interpretation of test results using a performance reference that corresponds to a lower age than that at which the student is tested.

This is distinct from modifying the test itself to enhance language comprehension, as seen in previous research. In our study, the test conditions and language-related requirements remain unchanged; the difference lies in how the test results are interpreted retrospectively. Nevertheless, the questions raised by Abedi, Hofstetter, and Lord regarding various accommodation strategies for testing newcomer students remain relevant to our examination of norm accommodation.

Our results show that norm accommodation paints a somewhat embellished picture of student outcomes that, at first misled us, researchers. We surmise that they could mislead parents or students themselves, as well as policy makers. Interviews conducted at the time of the research with schools and teachers from the reception classroom (Le Pichon et al. 2020), revealed that, indeed, teachers tend to norm-accommodate their students. This practice is used as a prevention strategy: Teachers want the students to feel at ease in the mainstream classroom as soon as they arrive and see the norm accommodation as an acceptable strategy.

A downside of this practice is that it may lower teachers' expectations by suggesting that the student actually functions on a lower age level, and consequently it may also lower self-esteem among the students, confirming the fears expressed by Abedi and colleagues in 2004, both self-esteem and teachers' expectations of their students being the keystones of achievement. A potential risk is that a practice like this one can become institutionalised as a disadvantage, where learners are assigned to different ability tracks on the basis of misrecognition of their competences. However, it is worth noting that certain newcomer students might be assigned to a grade level below their actual age, as mentioned in the work of Anisef et al. (2010, p. 122). In this situation, it would be inappropriate to use age as the standard reference point instead of grade level. This is because these students have not had the chance to acquire the academic knowledge and skills expected for the grade level that aligns with their chronological age.

Another intriguing finding of our research is the frequent norm accommodation observed in the results of non-newcomer students, or peers. This result can be attributed to the fact that newcomer students were enrolled in schools located in neighborhoods with a high concentration of migrant populations. It is not unlikely that within the classroom peer group, a significant number of students were also considered language learners to some extent. Regardless of whether norm accommodation should be considered as good practice or not, at present there appears to be no criteria for deciding that a student lacks proficiency in the language of the test and therefore needs accommodated norms, inevitably introducing a certain degree of arbitrariness. Given that a decision like this one is not trivial and can affect a student' s academic career over the long term, this raises ethical issues.

Adding to the complexity of the situation is the presence of early ability tracking systems, like the one employed in the Netherlands, which further exacerbates the

disadvantages faced by newcomer students (Field, Kuczera & Pont, 2007). Norm accommodation in this context poses the risk of conflating temporary academic delay with low academic level. Prior to entering secondary school (grades 5 and 6), students undergo centralized assessments without norm accommodation, producing results that may significantly deviate from previous academic performance, which had been influenced by varying degrees of norm accommodation. The resulting academic scores may pigeonhole learners into academic tracks that do not reflect their true academic potential, thereby raising concerns about the validity and reliability of such assessments (Peker, 2019). Moreover, we argue that norm accommodation can mask the extent of academic delay for some students, potentially delaying or impeding appropriate interventions to help them catch up with their peers.

In education, limited proficiency in the language of instruction can impede academic progress and lead to failure and early dropout. However, it is important to question whether the school language is the sole determining factor. What are the implications of our beliefs about the students themselves, our attitudes towards them, and the resulting practices (Rodriguez, 2018)? By interpreting student's performance at a lower age level than they are supposed to be evaluated at, we implicitly convey that they are incapable of performing at their age-appropriate level. While this may be justified for students who have had limited access to education, can this approach be indefinitely justified for all newcomer students or students with a migrant background? The risk lies in unintentionally creating a disadvantage for students by institutionalizing academic delay under the guise of equity.

7.2 What are the implications of this research for education policy?

Based on the results of this study, we advocate for more open discussions surrounding these practices, carefully evaluating their potential benefits, disadvantages, and impact on equity. It is crucial to involve parents, students, schools, and policymakers in this evaluation process. Notably, high-performing school systems prioritize equity and ensure that *all* students meet appropriate standards (Deng & Gopinathan, 2016). Abedi and colleagues' (2004) questions provide a valuable framework for reflection in this regard. Accommodations that correspond to the language needs of students (Bravo-Moreno, 2009), without underestimating their cognitive abilities, are essential. Numerous studies have examined the impact of translations, adaptations, multilingual dictionaries, software, and extra time on the performance of multilingual students (e.g., Attar, Blom & Le Pichon, 2020; Le Pichon et al., 2021).

Our study leads us to conclude that norm accommodation in educational systems is a practice driven by good intentions. However, it carries a potential drawback: it has the tendency to obscure the genuine academic requirements of students, lower the academic ambitions of both teachers and students, and mislead parents and policymakers regarding the academic accomplishments and progress of this vulnerable group of students.

Future studies may elucidate the prevalence and underlying motivations for the practice of norm accommodation by teachers. Testing accommodation practices, including norm accommodation, have the potential to be powerful tools for inclusion if implemented appropriately. An inclusive approach allows students not only to demonstrate their knowledge (Abedi Hofstetter & Lord, 2004), but also to develop according to age-related expectations thereby, enriching their own knowledge by continuing to develop in their own language alongside the school language. Newcomer students often have a rich but uneven and complex set of language and academic skills. The real challenge lies in validating these skills to remove barriers to vertical progression. This necessitates both cultural and structural reforms, including greater transparency in assessments and the establishment of a long-term academic vision shared with the students' families. Empowering tools must be implemented to enable newcomer students to effectively pursue their academic aspirations.

References

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: implications for policy-based empirical research. *Review of Educational Research*, 74(1), 1–28. https://doi.org/10.3102/00346543074001001

Anisef, P., Brown, R. S., Phythian, K., Sweet, R., & Walters, D. (2010). Early School Leaving among Immigrants in Toronto Secondary Schools. The Canadian Review of Sociology, 47(2), 103–128. https://doi.org/10.1111/j.1755-618X.2010.01226.x

Borodankova, O. & de Almeida Coutinho, A.S. (2011). *Grade Retention during Compulsory Education in Europe: Regulations and Statistics.* Brussels: Eurydice, p. 60. Available at: http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/126EN.pdf

Bravo-Moreno, A. (2009). Transnational mobilities: migrants and education. *Comparative Education*, 45(3), 419–433.

Browder, C.T. (2018). Recently Resettled Refugee Students Learning English in US High Schools : The Impact of Students' Educational Backgrounds. In Shapiro S., Farrelly, R. and Curry, M. J. (dir.), *Educating refugee-background students: Critical issues and dynamic contexts* (p. 17–32). Multilingual Matters.

Coelho, E., Oller, J., & Serra, J. M. (2011). Rethinking Initial Teacher Education for Linguistic and Cultural Diversity in the Classroom. *@ tic. revista d'innovació educativa*, (7), 52–61.

de Wal Pastoor, L. (2015). The mediational role of schools in supporting psychosocial transitions among unaccompanied young refugees upon resettlement in Norway. *International Journal for Academic Development*, 41, 245. doi:10.1016/j.ijedudev.2014.10.009

CITO (2019). *Toetsscore, vaardigheidsscore… en dan?* Brochure. Available at: https://www.cito.nl/-/media/files/ve-en-po/cito-flyer-toetsscore-vaardigheidsscore-en-dan.pdf

Collier, V. P., & Thomas, W. P. (1989). How quickly can immigrants become proficient in school English? *Journal of Educational Issues of Language Minority Students*, 5, 26–38.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada. A reassessment. *Applied Linguistics, 2*(2), 32–49.

Deng, Z. & Gopinathan, S. (2016). PISA and high-performing education systems: explaining Singapore's education success. *Comparative Education*, 52(4), 449–472.

Dryden-Peterson, S., Adelman, E., Bellino, M. J., & Chopra, V. (2019). The Purposes of Refugee Education: Policy and Practice of Including Refugees in National Education Systems. *Sociology of Education*, 92(4), 346–366. https://doi.org/10.1177/0038040719863054

Estrada, M., Burnett, M., Campbell, A., Campbell, P., Denetclaw, W., Gutiérrez, C., Hurtado, S., John, G., Matsui, J., McGee, R., Okpodu, C., Robinson, T., Summers, M., Werner-Washburne, M., & Zavala, M. (2016). Improving Underrepresented Minority Student Persistence in STEM. CBE Life Sciences Education, 15(3), es5–. https://doi.org/10.1187/cbe.16-01-0038

Feyisa, D., Mclean, C. & Lewis, K. (eds.). (2006). The Achievement of African Heritage Pupils: Good Practice in Lambeth Schools. London: Lambeth Research and Statistics Unit, p. 8, 18. Available at: http://www.lambeth.gov.uk/NR/rdonlyres/F3C58C99-F375-4679-B333-4F85D6CAC516/0/TheAchievementofAfricanHeritagePupils.pdf

Field, S., Kuczera, M. & Pont, B. (2007). No More Failures. Ten steps to equity in education. *Education and Training Policy*. OECD

Gagné, Schmidt, C., & Markus, P. (2017). Teaching about refugees: developing culturally responsive educators in contexts of politicised transnationalism. *Intercultural Education* (London, England), 28(5), 429–446. https://doi.org/10.1080/14675986.2017.1336409

Galton, M., Gray, J., & Ruddock, J. (1999). *The Impact of School Transitions and Transfers on Pupil Progress and Attainment.* Norwich: DfEE publications

Griebel, W., R. Heinisch, C. Kieferle, E. Röbe, & A. Seifert (Eds.) (2013). *Übergang in die Schule und Mehrsprachigkeit – Ein Curriculum für pädagogische Fach- und Lehrkräfte/Transition to School and Multilingualism – A Curriculum for Educational Professionals*. Verlag Dr. Kovač, Hamburg.

Gobbo, F., Ricucci, R. & Galloni, F. (2009). "Strategies for supporting schools and teachers in order to foster social inclusion: Italy". Inclusion and Education in European countries, p. 13. Available at: http://ec.europa.eu/education/more-information/doc/inclusion/italy_en.pdf

Hanushek, E.A. & Rivkin, S. (2010). "Generalisations about Using Value-Added Measures of Teacher Quality". Paper Presented at the annual meetings of the American Economic Association Atlanta, GA, January 3–5. Available at: http://www.usapr.org/paperpdfs/54.pdf

Haags Centrum voor Onderwijsadvies (2016). Tabellen tussenopbrengsten CITO LOVS 2016. Available at: https://dokumen.tips/documents/tabellen-tussenopbrengsten-cito-lovs-groepsniveau.html

Hollingworth, S. and Mansaray, A. (2012). Language Diversity and Attainment in English Secondary Schools: A Scoping Study; *The Institute for Policy Studies in Education* (IPSE) London Metropolitan University

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis*. Upper Saddle River, N.J: Pearson Prentice Hall.

Kieffer, M. J., & Thompson, K. D. (2018). Hidden progress of multilingual students on NAEP. Educational Researcher, 47(6), 391–398.

Koehler, C., N. Palaiologou & Brussino, O. (2022). Holistic refugee and newcomer education in Europe: Mapping, upscaling and institutionalising promising practices from Germany, Greece and the Netherlands, *OECD Education Working Papers*, No. 264, OECD Publishing, Paris, https://doi.org/10.1787/9ea58c54-en.

Koehler, C., Kakos, M., Sharma-Brymer, V., Schneider, J., Tudjman, T., Van den Heerik, A., Toumpoulidis, G. (2018). Multi-country partnership to enhance the education of refugee and asylum-seeking youth in Europe – PERAE (Comparative Report). *SIRIUS – Policy Network on Migrant Education*, Brussels. Retrieved from https://www.researchgate.net/publication/326113676_SIRIUS-Policy_Network_on_Migrant_Education_MULTI-COUNTRY_PARTNERSHIP_TO_ENHANCE_THE_EDUCATION_OF_REFUGEE_AND_ASYLUM-SEEKING_YOUTH_IN_EUROPE-PERAE_Comparative_Report

Koehler, & Schneider, J. (2019). Young refugees in education: the particular challenges of school systems in Europe. *Comparative Migration Studies*, 7(1), 1–20. https://doi.org/10.1186/s40878-019-0129-3

Le Pichon, E. (2020a). Bildung für Asylsuchende und Geflüchtete [Multilingualism in Society: Refuge and Asylum]. In. Gogolin, Hansen, McMonagle and Rauch. Handbuch "*Mehrsprachigkeit und Bildung*" [Handbook "Multilingualism and Language Education."] Springer. pp.245–250.

Le Pichon, E. (2020b). Intercultural communication, migration and mobility. In. Guido Rings & Sebastian Rasinger. *The Cambridge Handbook of Intercultural Communication*. Cambridge University Press handbook.

Le Pichon, E. & Baauw, S. (2020c). Advising municipalities on schooling Newly Arrived Migrant Pupils. In. Beerkens, R., Le Pichon, E., Supheerd, R., Ten Thije, J. eds. Enhancing Intercultural Communication in Organizations: Insights from Project Advisers. Routledge.

Le Pichon, E.; Cole, D.; Baauw, S.; Steffens, M.; van den Brink, M.; Dekker, S. (2020d). Transcultural itineraries and new literacies: how migration memories could reshape school systems, In. Passerini, Proglio and Trakilović, The Mobility of Memory across European Borders. Migrations and Diasporas in Europe and Beyond, Berghahn, New-York. Oxford.

OECD, 2010. "Closing the Gap for Immigrant Students: Policies, Practices and Performance". OECD Reviews of Migrant Education.

Peker, H. (2019). Assessment Considerations on English Learners and Their Disproportionate Placement in Special Education. *ELT Research Journal* 8(4), 221–235.

PISA, (2009). Results: Overcoming Social Background – Equity in Learning Opportunities and Outcomes (Volume II). Paris: OECD, 2010, p. 70. Available at: http://browse.oecdbookshop.org/oecd/pdfs/browseit/9810081E.PDF

Rodriguez, S. (2018). "Risky" Subjects: Theorizing Migration as Risk and Implications for Newcomers in Schools and Societies. *European Education*, 50(1), 6–26.

Schleicher, A. (ed.). (2003). Where immigrant students succeed – A comparative review of performance and engagement in PISA 2003. Paris: OECD, 2006, p. 111. Available at: http://www.oecd.org/dataoecd/2/38/36664934.pdf

Sprietsma, M. (2009). "Discrimination in Grading? Experimental Evidence from Primary School". Discussion Paper No. 09-074. Mannheim, Center for European Economic Research. Available at: http://ftp.zew.de/pub/zew-docs/dp/dp09074.pdf

Thompson, K. D., Umansky, I. M., & Porter, L. (2020). Examining Contexts of Reception for Newcomer Students. *Leadership and Policy in Schools*, 19(1), 10–35. https://doi.org/10.1080/15700763.2020.1712732

Traag, T. & van der Velden, R. (2011) Early school-leaving in the Netherlands: the role of family resources, school composition and background characteristics in early school-leaving in lower secondary education, *Irish Educational Studies*, 30:1, 45–62.

van de Werfhorst, H. (2018). Early tracking and socioeconomic inequality in academic achievement: Studying reforms in nine countries, *Research in Social Stratification and Mobility*, 58, 22–32. ISSN 0276–5624.

Volante, L., Klinger, D., Bilgili, Ö., & Siegel, M. (2017). Making sense of the performance (dis)advantage for immigrant students across Canada. *Canadian Journal of Education/Revue canadienne de l'éducation*, *40*(3), 229–361.