education sciences

MDPI

Article

Science Achievement of Multilingual Pupils: A Study on the Effectiveness of a Read-Aloud Assessment Accommodation

Fauve De Backer [1,*] and Lisa Dewulf [2]

[1] Department of Linguistics, Ghent University, 9000 Ghent, Belgium
[2] Department of Educational Sciences, Ghent University, 9000 Ghent, Belgium; lisa.dewulf@ugent.be
* Correspondence: fauve.debacker@ugent.be

Abstract: To date, empirical investigations of the effects of test accommodations on the actual achievement of multilingual pupils have been inconclusive. In this present study, we investigated whether read-aloud accommodation contributes to better results in terms of science achievement for multilingual pupils. A computer-based science test, conducted with or without read-aloud accommodation, was administered to 1022 5th-grade pupils in 36 Flemish primary schools. We assessed the hypotheses that, first, pupils in a condition with accommodation perform better than their non-accommodated peers, and second, certain background characteristics are related to science achievement for different groups of pupils. The results indicate that read-aloud accommodation in language education does not significantly contribute to making assessments fairer. Overall, parental job status, grade retention, migration status, and self-reported oral proficiency significantly predicted pupils' science achievement. For pupils taking an accommodated test, their age of arrival and the language they spoke at home did not significantly relate to their science achievement, but their self-rated literacy skills in the language of schooling did.

Keywords: assessment; testing; evaluation; accommodations; read-aloud protocols; multilingualism; effectiveness; validity; primary education

check for updates

Citation: De Backer, F.; Dewulf, L. Science Achievement of Multilingual Pupils: A Study on the Effectiveness of a Read-Aloud Assessment Accommodation. *Educ. Sci.* **2024**, *14*, 494. https://doi.org/10.3390/educsci14050494

Academic Editor: Lawrence Jun Zhang

Received: 25 January 2024
Revised: 1 May 2024
Accepted: 2 May 2024
Published: 4 May 2024

1. Introduction

There is an increasing concern that multilingual pupils (MPs) are disadvantaged in comparison to L1 speakers (L1S) of the language of schooling in traditional testing practices since their linguistic ability affects their results on content-related tests [1–3]. In content-based subjects such as mathematics or science, the goal of tests is to assess what a pupil knows and can do concerning these subjects. Language factors should not confound test results, but every test that uses language cannot avoid also being a language proficiency test [4]. The goal cannot be to completely eliminate the impact of language factors on test results. It is much more a matter of reducing construct-irrelevant factors. To ensure fair and valid tests, test accommodations have been proposed [5]. In the present study, we investigated whether read-aloud accommodation contributes to better results in science achievement for MPs. According to the Council of Europe [6], multilingualism is understood as the coexistence of multiple languages within a given society. Given the diverse linguistic landscape and proficiency levels of individuals in our research context, we adopt the term 'multilingual pupils' to encompass individuals at all points along the linguistic continuum.

1.1. Assessment Accommodations

Widely varying definitions of test accommodations have emerged. In the present study, we adopted the definition of Butler and Stevens [7]: "Support provided to students for a given testing event either through modification of the test itself or through modification of the testing procedure to help students access the content in English and better demonstrate what they know" (p. 5). For MPs, linguistic support accommodations such as multilingual tests, dictionaries or glossaries, language simplifications or read-aloud protocols are suggested, but research into how effective these accommodations are and under what conditions is limited [8]. Their effectiveness is most often evaluated in terms of increases in pupils' test results and decreases in the impact of construct-irrelevant factors.

1.1.1. Read-Aloud Accommodations

Concerning read-aloud accommodations, the research field is rather young and indecisive, which underlines the necessity to expand it. Pupils have different preferences: some students might be more prone to take a test with visual and oral stimuli than a test with only visual stimuli. For example, pupils who do not understand a test question might benefit from having the words read aloud with the proper intonation [9]. Such accommodations are classified by Rivera and colleagues [10] as some of the most effective supports, a claim that is supported by Kieffer and colleagues, who state that reading test questions aloud may address the needs of MPs [11]. However, in their meta-analysis, Pennock-Romand and Rivera [12] included 14 US studies with English language learners to test accommodation versus control conditions. The read-aloud accommodation was based on only one study [13], where it showed zero effects on its own, and in combination with a pop-up dictionary. Reed and colleagues [14] investigated the relative effects of teachers' read-alouds and pupils' silent reading in a randomized controlled trial. Their results show that pupils can learn and retain information and content-based vocabulary equally well when they read the informational text silently themselves compared to when the teacher reads the text to them. This indicates that read-aloud accommodations might not be as effective as was suggested by Rivera and colleagues [10]. Buzick and Stone [15] compared the effects for students with and without (learning) disabilities and concluded that this kind of accommodation is beneficial for both groups in terms of increasing their reading scores, but that it yields only small gains for mathematics scores. This suggests that the read-aloud results might be beneficial for pupils in some subjects, but not for all. And finally, Abedi [1] reports how read-alouds have been used for pupils who are in the process of acquiring the language of schooling, without any indication of the relevance or effectiveness.

1.1.2. Explaining Pupils' Differences

In sum, the research into the effectiveness of read-aloud accommodations seems rather contradictory and many issues remain. Attention has only been given recently to research into accommodations for MPs, and only a limited number of studies on read-aloud accommodations have controlled for pupil background variables [16]. It could be that these accommodations are beneficial for some pupils under certain conditions, but not for others. The fact that MPs are not a homogeneous group [17] is becoming more widely acknowledged when considering accommodations. According to Butler and Stevens [7], determining which background variables are most indicative of student performance may help assessment specialists to decide which accommodations would be most advantageous to specific pupils.

Although research in this area is limited, several studies have shed light on the influence of background characteristics on academic achievement. For example, gender, grade retention, and socioeconomic status (SES) have been found to be associated with science achievement. The impact of gender on science performance remains somewhat ambiguous. While findings from the Trends in International Mathematics and Science Study (TIMSS) suggest that boys in the fourth grade in Flanders tend to outperform girls in science [18], results from the Programme for International Student Assessment (PISA) 2015 show comparable science performance levels between genders [19]. Numerous studies have explored the potential effects of grade retention, with one study indicating that, specifically for multilingual pupils, grade retention initially correlates negatively with science achievement. However, this association becomes non-significant when considering proficiency in the language of instruction and reading performance [20]. SES has consistently been linked to academic achievement, including science performance. Belgium stands out among high-performing countries in terms of the strength of the relationship between SES and student performance [21]. Regarding migration status, PISA results indicate the existence of an average difference in science performance between immigrant and non-immigrant students, even after adjusting for SES [21]. Additionally, older children who immigrate tend to perform less well on reading assessments at age 15, with late arrival and limited language proficiency in the schooling language exacerbating vulnerability [22].

1.2. Research Aim

The effectiveness of a read-aloud accommodation was explored through a randomized experiment. Both pupils who were not yet proficient in the language of schooling and therefore needed the accommodations and pupils who did not need the accommodations were randomly assigned to different conditions. If pupils who did not need the accommodation performed significantly differently when assigned the accommodation, this would provide evidence that the accommodation changed the focal construct [1]. A unique aspect of this study was the documentation of the frequency with which students used the read-aloud accommodation. Additionally, we aimed to contribute to the existing knowledge by presenting results from a different context—specifically, Flanders, the Dutch-speaking region of Belgium. In this field of study, research on accommodation is predominantly conducted in the United States [1].

Due to Belgium's migration history, children of non-Western European descent primarily trace their roots to Turkish, Moroccan, or Eastern European origins. Furthermore, those with Turkish or Moroccan ancestry often belong to second or third generations, while individuals with Eastern European backgrounds are predominantly first-generation immigrants. Consequently, the Flemish context offers a research environment rich in linguistic diversity, where both the language spoken at home and the proficiency in the language of instruction (Dutch in Flanders) vary depending on the family's migration history. This diversity provides a distinctive opportunity for evaluating the efficacy of accommodations across a broad spectrum of linguistic backgrounds. It is within this context that Turkish and Polish pupils were specifically selected for inclusion in this study.

We assessed the hypotheses that, first, pupils in a condition with accommodation perform better than their non-accommodated peers, and second, that certain background characteristics are related to science achievement. The analyses were first carried out for the complete group of pupils who participated in this study. They were then replicated for only the MPs who were in the accommodated test condition. We took these actions in order to compare differences in how background characteristics relate to science achievement.

2. Materials and Methods

2.1. Participants

Data were collected from 1022 fifth-grade pupils (aged 10–11 years, 49.3% boys, 50.7% girls) in 36 primary schools in Flanders. We selected schools in urban regions in Flanders using a proportional criterion of pupils speaking a language at home other than the language of schooling. If pupils did not exclusively speak the language of schooling at home, they were identified as MPs (87.6% in the present sample). Among the MPs, there was not one common home language, and the pupils varied widely in terms of linguistic backgrounds. The age at which pupils arrived ranged from born in the host country (68% of the pupils) to 12 years, with an average arrival age of 1.75 years (SD = 3.04) and with few participants arriving at age 12 ($n = 2$) or 11 ($n = 14$).

2.2. Procedure

The data were collected in two rounds: the first half of the school year 2016–2017 and the first half of the school year 2017–2018. Participating schools were visited by the first author, who tested all fifth-grade pupils during regular class periods. Pupils were randomly assigned to one of the two research conditions: they either took an accommodated ($n = 515$) or a non-accommodated ($n = 507$) computer-based science test.

The focus of this present study is on the conditions with read-aloud accommodations: the non-accommodated test or the 'Dutch without read-alouds test' (DU/A-), and the test with read-alouds in Dutch (DU/A+). In conditions with read-aloud support, the pupils were offered the possibility to listen to a read-aloud version of each question and the multiple-choice answers. Pupils filled in a background questionnaire and took the science test on a computer. Read-aloud accommodations in assessment can take different forms. The test administrator or teacher can read the questions out loud in the language of schooling, or a translator can read aloud a previously translated script in the mother tongue of the pupil (scripted oral translation), or the test items can be recorded beforehand. This last option guarantees a standardized test administration, which has the advantage of ensuring that there are no differences in, for example, pronunciation, volume, or pace [23]. However, one disadvantage of recorded accommodations is that learners do not have the visual cue of seeing the person's lips/mouth move, which can be beneficial for understanding, especially for multilingual learners. Because of these advantages, we opted for recorded accommodations. In this way, we were able to minimize variation in test administration, which can create construct-irrelevant variance [16]. We opted for voicing by an L1 speaker with standard pronunciation and intonation patterns. Video instructions on how to take the test were developed for each condition. Pupils were not given any time constraints, which is innovative because read-aloud accommodations have not yet been studied with generous time conditions [12].

2.3. Variables

Dependent Variables. Science achievement was measured by means of the TIMSS 2011 released science items, which can be considered a standardized, curriculum-based achievement test (Cronbach's alpha of the present sample: 0.71). This was chosen to maximize equivalence between the original and the translated versions. The test consisted of the 43 items and had a multiple-choice format. The test was administered in the first semester of the school year.

Independent Variables. A pupil background questionnaire and information about accommodations and groups were used to gather data about the pupils. Five clusters of pupil information were studied:

(1) We assessed background characteristics. Pupil information about gender (boy = 0, girl = 1) and grade retention (0 = on track, 1 = grade retention) was gathered. Parental job status was measured based on the highest parental job status of either parent, and pupils were asked about their professional occupations. Answers were coded on the basis of the scheme provided by Erikson, Goldthorpe, and Portocarero [24]. To determine the parental job status (categorical variable), only the highest the parental job status of parents was taken into account in the analyses. For instance, if the mother of a pupil X is a professional and the father of pupil X a specialized manual worker, the highest parental job status of the pupil X is coded as 8. The scale used was as follows: 0 = inactive on the labour market (unemployed, retired, houseman/-wife, sickness, etc.), 1 = unskilled manual workers, 2 = specialized manual workers, 3 = skilled manual workers, 4 = routine non-manual employees, 5 = self-employed and small proprietors, 6 = lower grade employees and administrators, 7 = higher-grade administrators and executives, 8 = professionals, entrepreneurs, and large proprietors. The status of inactive on the labour market was selected as the reference category.

(2) Oral language proficiency was measured using a scale in which pupils had to self-assess the extent to which they were proficient in speaking and listening in both the language of schooling (5 items) and their L1 (2 items) on a five-point Likert scale (1 = very poor—5 = very well) (respectively, Cronbach's alpha = 0.74 and 0.83; CR = 0.83 and 0.92). Also, we studied the language spoken at home when MPs were with their fathers (1 = always one or more languages other than Dutch,

2 = often another language or more languages than Dutch, 3 = sometimes Dutch, sometimes other language(s), 4 = most of the time Dutch, and 5 = always Dutch). Never Dutch (= 1) was chosen as the reference category. Concerning language use in the home, De Houwer [25] that differences in language input by the parents predicted the success rate of raising actively multilingual children. We chose language use with the father because Duursma and colleagues [26] explored the impact of language preference with the father on the language abilities of pupils, and it appeared that language choice with MPs' fathers was a significant predictor, not only of the language competences in the home language but also in the school language. Moreover, pupils' language use with their fathers also appears to be a crucial predictor of science achievement, rather than the language pupils use with their mother or their siblings [27].

(3) We investigated cultural proximity. The migration status of pupils (0 = non-immigrant [= reference category]; 1 = first generation—pupil is foreign-born; 2 = second generation—pupil is native-born but both parents are foreign-born; 3 = third generation—pupil and parents are native-born, but at least one grandparent is foreign-born; and 4 = 2.5 generation—1 of pupils' parents are foreign-born) and age of arrival (in Belgium) were questioned.

(4) We assessed the level of literacy. Pupils' were asked to self-assess literacy proficiency in Dutch (scale consisting of 5 items, Cronbach's α = 0.797; CR = 0.69) and their literacy proficiency in L1 (2 items, Cronbach's α = 0.89; CR = 0.95) on a five-point Likert scale (1 = very bad—5 = very good), which included items such as 'How well can you read in Dutch' or 'Writing an e-mail in Dutch is easy' (1 = I totally disagree, 2 = I disagree, 3 = in between, 4 = I agree, 5 = I totally agree).

(5) We assessed the frequency of use. Pupils had the opportunity to listen recordings as often as they wanted to. The proportion of clicks on the audio buttons (clicks/total available audio) was calculated and this was used as a variable for measuring the frequency at which the read-aloud accommodation (M = 0.08; SD = 0.12) is used.

The average use of the available audio support was rather low. Figure 1 provides an overview of the pupils' use. Only some pupils used the audio, and they did not do this frequently. The majority of pupils used it occasionally. None of the pupils used it for every question. The frequency of use was similar for MPs and L1S.

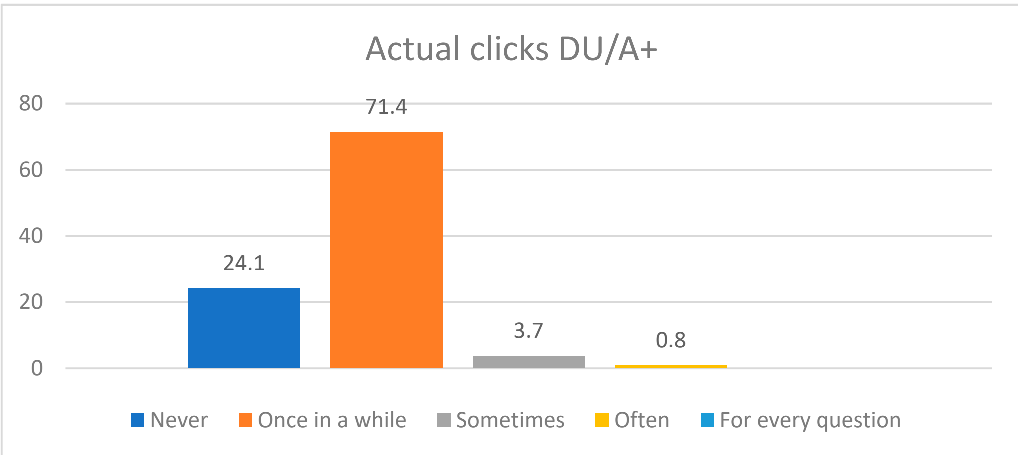

Figure 1. Frequency of use of all accommodated pupils—by categories.

Table 1 provides an overview of the correlates and descriptive statistics for all pupils. Table 2 provides an overview of the descriptive statistics of relevant variables for L1S of the language of schooling and MPs separately.

Table 1. Correlates and descriptive statistics.

	M (SE)	1	2	3	4	5	6	7	8	9	10	11	12
1. Science Achievement	18.38 (0.30)	1	0.169 **	0.062 *	−0.144 **	0.217 **	−0.073 *	0.200 **	−0.082 **	−0.120 **	0.207 **	−0.065 *	0.016
2. Parental job status	2.21 (0.61)		1	0.027	−0.099 **	0.044	−0.014	0.156 **	0.034	−0.068 *	0.067 *	−0.012	−0.009
3. Gender	-			1	0.014	0.024	0.046	−0.037	−0.031	0.006	−0.036	−0.009	0.000
4. Grade retention	0.32 (0.02)				1	−0.55	0.055	−0.083 **	0.008	0.109 **	−0.068 *	0.012	−0.005
5. ProfDutch	4.27 (0.02)					1	0.118 **	0.207 **	0.026	−0.167 **	0.676 **	0.051	0.028
6. ProfL1	4.32 (0.03)						1	−0.212 **	0.001	0.092 **	0.075 *	0.505 **	−0.009
7. LangFath	2.63 (0.04)							1	0.001	−0.211 **	0.221 **	−0.025	0.044
8. Migstat	1.86 (0.04)								1	−0.405 **	0.102 **	−0.005	−0.036
9. Age of arrival	1.75 (0.10)									1	−0.267 **	0.102 **	0.020
10. LitDutch	4.37 (0.20)										1	0.015	0.016
11. LitL1	3.35 (0.04)											1	0.009
12. Frequency of use	0.08 (0.12)												1

* $p < 0.05$, ** $p < 0.01$.

Table 2. Descriptive statistics for Dutch language (DL) speakers and multilingual pupils (MPs).

	M (SE) DL	Min DL	Max DL	M (SE) MP	Min MP	Max MP
1. Science Achievement	21.03 (5.94)	2	37	18.36 (5.18)	0	38
2. Parental job status	3.12 (2.36)	0	8	2.12 (1.82)	0	8
4. Grade retention	0.27 (0.45)	0	1	0.33 (0.47)	0	1
5. ProfDutch	4.4 (0.51)	2.4	5	4.26 (0.55)	2.2	5
10. LitDutch	4.55 (0.50)	2.4	5	4.36 (0.61)	2	5

2.4. Data Analysis

The data used in this study have a hierarchical structure: 1022 pupils are nested within 36 schools. To test (A) the relationship between science achievement and the type of accommodations pupils are receiving and (B) the relationship of pupil characteristics with science achievement, we applied multilevel modeling based on hierarchical regression. We conducted this using MlwiN 2.29. This multilevel analysis comprised a two-step procedure and was used to determine the group effect. First, a two-level null model was estimated to analyze the raw science achievement of pupils. The two-level null model (Model 0) was utilized to investigate whether a multilevel approach was preferred over single-level regression analysis. Second, information about the accommodation was added to the model (Model G).

A multi-step procedure was used to determine the effect of the pupil predictor variables. First, the two-level null model was estimated (Model 0). In the second and the following steps (Models 1–4), a two-level random-intercepts model was estimated. By using a stepwise multilevel approach, the additional value of each subset of variables to the model was checked. Before the next subset of variables was added to the model, the non-significant factors were deleted in order to obtain the most parsimonious model possible.

3. Results

3.1. Assessment Accommodations

Model 0. The fully unconditional two-level null random-intercepts model (Model 0) predicts the overall score for science achievement for all fifth-grade pupils (Level 1) across the schools (Level 2). The intercept of the null model represents the overall mean score for science achievement for all pupils across all schools (M = 18.38).

The null model divides the science achievement scores based on whether variance occurs due to differences pupils or between schools. The results state that the proportion of variance due to differences between schools was 7%, while 93% of the total variance was situated at the pupil level. As illustrated in Table 3, the variances in science achievement scores on the two levels, pupil (χ^2 (1) = 7.51, $p < 0.01$) and school (χ^2 (1) = 492.84, $p < 0.001$), were significantly different from zero. This justifies the use of multilevel modeling for studying pupils' achievements in terms of science achievement.

Table 3. Multilevel parameter estimates for the two-level analysis: all pupils.

	Model 0		**Model 1**		**Model 2**		**Model 3**	
	Coef.	***SE***	**Coef.**	***SE***	**Coef.**	***SE***	**Coef.**	***SE***
Fixed parameters								
Intercept	18.379 ***	0.300	17.752 ***	0.427	17.611 ***	0.482	21.072 ***	0.896
Job status								
(ref: no job)								
Unskilled manual workers			0.939	0.492	0.887	0.499	0.701	0.505
Specialized manual workers			2.097 ***	0.549	1.756 **	0.549	1.526 **	0.555
Skilled manual workers			0.854	0.496	0.660	0.493	0.545	0.504
Routine non-manual employees			2.195 ***	0.584	1.781 **	0.585	1.833 **	0.596
Self-employed and small proprietors			−0.592	1.006	−0.864	0.991	−1.050	0.973
Lower grade employees and administrators			2.881 ***	0.761	1.940 *	0.771	1.115	0.774
Higher-grade administrators and executives			4.929 **	1.626	3.089	1.757	2.359	1.722
Professionals, entrepreneurs and large proprietors			3.878 *	1.619	3.735 *	.	3.587 *	1.624
Grade retention			−1.366 ***	0.351	−1.292 ***	0.349	−1.273 ***	0.353
(ref: no grade retention)								
Proficiency Dutch					1.750 ***	0.306	1.688 ***	0.312
Language use with father								
(ref: never Dutch)								
Mostly other language(s) than Dutch					−0.001	0.531	0.087	0.543
Sometimes Dutch, sometimes other language(s)					−0.055	0.411	−0.137	0.418
Mostly Dutch					1.865 **	0.615	1.924 **	0.627
Always Dutch					1.680 **	0.548	0.548	0.612
Migstatus								
(ref: no migration status)								
First generation							−2.078 *	0.960
Second generation							−3.818 ***	0.784
Third generation							−5.825 ***	1.278
2.5 generation							−3.628 ***	0.791
Age of arrival							−0.252 **	0.095
Random part								
Level 2: school	2.040 **	0.744	1.614 *		1.545 *	0.615	1.423 *	0.586
Level 1: pupil	26.524 ***	1.195	24.566 ***		22.726 ***	1.089	21.564 ***	1.068
Model fit								
Deviance (-2LL)	6276.794		5873.887		5422.218		5047.906	

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.0$.

Model G. Information regarding groups was added to the model. We selected MPs that were not accommodated (MP-nonaccommodated) as the reference category. The results show that the group accommodation pupils were assigned to was not significantly related to differences in the science achievement of the pupils (MPaccommodated: $B = 0.380$, χ^2 (1) = 1.25, $p > 0.05$; L1Saccommodated: $B = 2.248$, χ^2 (1) = 7.70, $p < 0.01$; L1Snonaccommodated: $B = 3.03$, χ^2 (1) = 13.86, $p < 0.001$). Read-aloud accommodations were not related to science achievement, but L1S had significant better science results than MPs.

3.2. All Pupils

Model 1. First, pupil background characteristics (Level 1) were included in a stepwise manner in the model. Gender, parental job status, and information about grade retention were added to the fixed part of the model to predict the score in terms of science achievement. Since gender was not significant in this model, it was omitted in Model 1 in order to obtain the most parsimonious model possible. In Model 1, we observed a positive fixed slope for the parental job status for specialized manual work ($B = 2.10$, $SE = 0.55$, $df = 1$, $p < 0.001$); routine non-manual work ($B = 2.20$, $SE = 0.58$, $df = 1$, $p < 0.001$); lower middle management (B = 2.88, $SE = 0.76$, $df = 1$, $p < 0.001$); higher middle management ($B = 4.93$, $SE = 1.63$, $df = 1$, $p < 0.01$); and liberal professions, top managers, top professionals, and administrators ($B = 3.88$, $SE =1.62$, $df = 1$, $p < 0.05$). Grade retention was negatively related to science achievement ($B = -1.37$, $SE = 0.35$, $df = 1$, $p < 0.001$). On the basis of a comparison of the deviance, Model 1 fitted the data better than Model 0 (χ^2 (9) = 402.91, $p < 0.001$).

Model 2. Next, the average level of proficiency in Dutch, the level of proficiency in L1, and the language spoken at home with the father were added to Model 1. Parental job status and grade retention were still significantly related to science achievement. The proficiency level in Dutch and the language spoken with the father at home were significantly related to science achievement. The proficiency in L1 was not significantly related to science achievement and it was removed from the model. In Model 2, pupils with a better proficiency level in Dutch had better science results ($B = 1.75$, $SE = 0.31$, $df = 1$, $p < 0.001$). Pupils who spoke mostly Dutch with their father at home had better science results ('mostly Dutch': $B = 1.87$, $SE = 0.62$, $df = 1$, $p < 0.01$). Model 2 fitted the data significantly better than Model 1 (χ^2 (5) = 451.67, $p < 0.001$).

Model 3. Next, the migration status and the average age of arrival in Belgium were added. Both variables were significantly related to science achievement. In Model 3, having a 1st-, 2nd-, 2.5th-, and 3rd-generation migration status was negatively related to the science achievement of pupils (3rd generation $B = -5.83$, $SE = 1.28$, $df = 1$, $p < 0.001$). The later a pupil arrived in Belgium, the lower the science achievement was ($B = -0.25$, $SE = 0.10$, $df = 1$, $p < 0.001$). The age of arrival was significantly and negatively related to science achievement. Model 3 fitted the data significantly better than Model 2 (χ^2 (5) = 374.31, $p < 0.001$).

Model 3a. In Model 3a, self-perceived literacy skills in Dutch and L1 were added. None of these parameters were significantly related to science achievement. Consequently, they were omitted and Model 3a is not reported in the table.

Model 3b. In Model 3b, we added the average use of the accommodations, the number of which was not significantly related to science achievement. Consequently, Model 3a was considered as the final and most parsimonious model and Model 3b is not reported in the table.

This stepwise modeling was repeated three times for the different conditions (all pupils, MPs, and MPs in the accommodated condition).

For all pupils, 19.52% of the variance in the science achievement test was explained by the predictors shown in Model 3.

3.3. Multilingual Pupils in the Accommodated Condition

The stepwise, multilevel approach—analogous to the analysis for all pupils and MPs—was conducted for the MPs in the accommodated condition as well. For the MPs in the accommodated condition, only the null and final Model 4 are reported since information about the frequency of use was not significant and therefore left out. The intermediate steps are reported in Table 4

The total variance of the science scores was 26.25, which was the sum of the school (Level 2) variance (1.18) and the pupil (Level 1) variance (25.07). The results state that the proportion of variance due to differences between schools was 4%, and that 96% of the total variance was situated at the pupil level. As illustrated in Table 4, the variances in the science achievement scores at the pupil level (χ^2 (1) = 2.32, $p > 0.05$) were not significantly different from zero.

In the final Model 4, parental job status and grade retention were significantly related to science achievement. Pupils with a father or mother who was a routine non-manual worker performed significantly better in terms of science achievement, while grade retention was negatively related to the science achievement of MPs with the accommodated condition. The proficiency level in Dutch was significantly related to science achievement, while the language spoken at home was not. A better proficiency level in Dutch resulted in better science results, except when the frequency of which pupils' use the read-alouds was added as a categorical variable (see 'Model extra' in Table 4).

Being 2nd-, 2.5th-, and 3rd-generation was significantly and negatively related to science achievement. The age of arrival was not significantly related to science achievement. Finally, for MPs in the accommodated condition, a higher level of literacy skills in Dutch was positively related to science achievement.

For MPs in the accommodated condition, 19.51% of the variance in the science achievement test was explained by the predictors in Model 4.

Table 4. Multilevel parameter estimates for the two-level analysis: multilingual pupils in the accommodated condition.

	Model 0		**Model 1**		**Model 2**		**Model 3**		**Model 4**		**Model Extra**	
	Coef.	**SE**	**Coef.**	**SE**	**Coef.**	**SE**	**Coef.**	**SE**	**Coef**	**SE**	**Coef**	**SE**
Fixed parameters												
Intercept	18.475 ***	0.310	18.131 ***	0.551	18.399 ***	0.546	21.277 ***	1.151	21.468 ***	1.191	21.480 ***	1.393
Job status *(ref: no job)*												
Unskilled manual workers			0.862	0.714	0.388	0.711	0.586	0.705	0.665	0.710	0.997	0.789
Specialized manual workers			2.017 *	0.799	1.712 *	0.797	1.570 *	0.795	1.496	0.796	1.580	0.856
Skilled manual workers			0.704	0.712	0.398	0.703	0.564	0.703	0.492	0.705	0.634	0.765
Routine non-manual employees			2.683 **	0.861	2.602 **	0.853	2.814 **	0.858	2.753 **	0.859	3.456 **	0.945
Self-employed and small proprietors			−0.926	1.552	−1.062	1.512	−1.060	1.483	−1.261	1.477	−0.129	1.534
Lower grade employees and administrators			1.489	1.122	0.558	1.153	−0.082	1.167	−0.031	1.195	0.914	1.315
Higher-grade administrators and executives			5.432*	2.248	4.365	2.438	4.501	2.393	4.297	2.379	5.226 *	2.350
Vrije beroepen, Professionals, entrepreneurs and large proprietors			2.177	3.487	2.052	4.774	−0.703	4.775	−1.007	4.749	−0.684	4.670
Grade retention (ref: no grade retention)			−1.761 ***	0.510	−1.817 ***	0.505	−1.865 ***	0.505	−1.861 ***	0.507	−1.715 ***	0.563
Proficiency in Dutch					2.192 ***	0.441	2.208 ***	0.437	1.196 *	0.575	0.915	0.610
Migstatus *(ref: no migstatus)*												
First generation							−2.053	1.111	−2.078	1.140	−2.414 *	1.195
Second generation							−3.646 **	1.106	−3.928 ***	1.130	−4.143 ***	1.169
Third generation							−3.621	1.932	−3.822 *	1.934	−4.693 *	2.027
2.5 generation							−3.300 **	1.174	−3.560 **	1.198	−3.719 **	1.264
Literacy in Dutch									1.489 **	0.520	1.332 *	0.557
Use of read alouds *(ref: never)*												
Little											−0.117	0.700
Occasionally											−0.696	0.809
Sometimes/Often											0.956	1.269
Random part												
Level 2: school	1.179	0.775	0.733	0.639	0.670	0.616	0.387	0.528	0.435	0.540	0.000	0.000
Level 1: student	25.073 ***	1.735	23.189 ***	1.651	21.886 ***	1.599	21.056 ***	1.561	20.696 ***	1.547	20.050 ***	1.603

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.0$.

4. Discussion

The findings indicate that the performance gap persists between pupils who speak the language of schooling at home and those who do not, even when considering socioeconomic status (SES). Previous research has shown that MPs with the same ability levels as L1 speakers do not have the same probability to solve a test item correctly [28] because their linguistic abilities impact the results [2]. To address this issue with test validity, accommodations are recommended.

As this research field has only recently begun to emerge and as there is substantial demand for experimental studies [29], our goal is to make a contribution to the understanding of assessment accommodations. We examined whether read-aloud protocols in the language of schooling can succeed in leveling the playing field for multilingual learners. We opted for read-aloud accommodations because they are commonly used by teachers in Flanders, reflecting their acceptance and integration into educational practices. Unlike some other accommodations, there is little controversy surrounding the use of read-aloud accommodations, making them a suitable focus for our investigation into the effectiveness of supporting multilingual pupils during assessments.

The results of the first research question, whether pupils in a condition with read-aloud accommodation perform better than their non-accommodated peers, do not show the existence of a significant relationship between science achievement and read-aloud accommodation. As Castellon-Wellington [9] points out, a possible explanation for the absence of significant effects on pupils' performance is that pupils may not be familiar with the academic vocabulary used in the test. In cases where the vocabulary is unknown to the pupil, read-alouds may not be the most suitable accommodation for addressing the needs of MPs. According to the simple view of reading [30], two interrelated components that develop at the same time are necessary: decoding and language comprehension. Accommodating for one—in this case, decoding—may not have the intended benefit of boosting language comprehension. In other words, compensation for decoding difficulties may improve reading comprehension, but only if language comprehension is already adequate [14]. Cohen and colleagues [31] also suggest that the effects of these interventions may be different when accommodations are explored for older pupils compared to younger pupils, and that more research is needed in terms of cognitive load. Indeed, there are more cognitive resources required in a test with both visual and oral "distractions".

The frequency of use was not a significant predictor of science achievement. However, it appeared that the average frequency with which the read-aloud accommodation in the language of schooling (M = 0.08; SD = 0.12) was used was not particularly high, which was possibly one of the explanations as to why we did not find a significant difference between groups. The use of video instructions to enable pupils to make use of the accommodations made may not have been sufficient. The present study seems to contradict the suggestions of Acosta and colleagues regarding read-alouds[16], who suggest that experience with the accommodation prior to the assessment is not strictly necessary, as opposed to other accommodations such as a dictionary.

The second question focused on whether there are different background characteristics related to science achievement, both for the entire group of pupils (n = 1022) and for MPs who received the read-aloud accommodation (n = 450). Pupils from families with higher job status also achieved higher results in science achievement [21]. Grade retention was negatively related to science achievement, which is supported by previous findings [32]. We did not find any significant differences between boys and girls in terms of science achievement, which was similar to the PISA 2015 results [19]. Also in line with the PISA data, our results indicated a difference in science performance between immigrant and non-immigrant pupils [33]. Pupils' self-reported oral proficiency in the language of schooling was significantly related to science achievement for all three groups analyzed, which was similar to the meta-analysis of Prevoo and colleagues [34], who found that MPs who are more orally proficient generally have better school outcomes.

The age of arrival was a significant predictor of science achievement for the entire group, but it did not explain any extra variance when we only considered accommodated MPs. In other words, when MPs are accommodated for a test, the age at which they arrived in the host country does not significantly predict their science achievement anymore.

The frequency with which pupils speak the language of schooling with their father appeared a significant predictor of science achievement for pupils, but not when only the accommodated group was accounted for. Thus, it seems that when pupils are not accommodated for a test, the less they speak the language of schooling with their fathers, the lower their performances on the science test will be. Interestingly, we did not observe this negative relationship when pupils were accommodated for. One possible explanation for this is that accommodating for a test moderates the relationship between language use with the father and science achievement.

Research has already revealed the importance of literacy skills in the language of schooling for science achievement [35,36]. In the present study, self-reported literacy skills were not a significant predictor when the analyses were performed for all pupils. Surprisingly, however, it did add to the explained variance when only the accommodated MPs were taken into consideration. Thus, when pupils are accommodated for a science test with read-aloud support in the language of schooling, their self-reported literacy skills are positively related to their science achievement. More research is needed to explain this finding, but these results confirm the importance of literacy [37], a life skill which may be also needed to make effective use of the accommodations provided.

This exploratory study has several limitations that should be taken into account. First, the fact that the sample size of the pupils who are L1 speakers of the language of schooling consisted of only 89 pupils should not escape our attention. Also, we used data reported by the pupils for language proficiency and literacy: we did not want to burden these young pupils with too many tests. While it is common practice in educational research to rely on self-reported data, their accuracy is still open for debate. The meta-study of Kuncel, Credé and Thomas [38] indicated that self-reported data are generally accurate and can be used as a measure of pupil achievement. However critical voices must be notedtoo. For example, Rosen, Porten, and Rogers [39] caution that lower-performing pupils in aprticular tend to overestimate their grades.

5. Conclusions

In this study, our findings indicated that read-aloud accommodations in the language of schooling did not lead to fairer assessments, aligning with previous research that highlighted the significant performance gap between first-language speakers and multilingual pupils. It is worth noting that while read-aloud accommodations may offer advantages, other accommodations may be more adept at addressing the language barrier, indicating that the frequency of use by teachers and the absence of controversy should not be the sole criteria for determining their suitability in supporting multilingual pupils during assessments. These results underscore the importance of exploring alternative accommodations that specifically address the language-related challenges faced by multilingual learners. While we recognize that we cannot alter learners' backgrounds, our study sheds light on the potential effectiveness of tailoring accommodations to mitigate language barriers.

Moving forward, it is crucial to consider accommodations that target the specific linguistic needs of multilingual pupils, especially in assessments where language proficiency plays a significant role. Exploring accommodations that go beyond read-aloud protocols and directly address the challenges associated with second language acquisition could lead to more equitable assessment practices. Additionally, our findings emphasize the importance of considering various pupil characteristics, such as parental job status, grade retention, migration status, and self-reported oral proficiency, in order to understand their science achievement levels.

By delving into the relationship between pupil characteristics and science achievement, both within the complete group of pupils and among those receiving read-aloud accommodations, our study provides valuable insights into the complex interplay between background factors and academic performance. Future research endeavors should focus on further investigating the effectiveness of accommodations that are tailored to the needs of multilingual learners, with the aim of promoting fairer and more inclusive assessment practices in educational settings.

Author Contributions: Conceptualization, project administration, methodology, investigation, writing, reviewing and editing, F.D.B. Analysis, L.D. All authors have read and agreed to the published version of the manuscript.

Funding: The research reported in this article was funded by the Flemish Research Foundation [grant number FWOOPR2015003901].

Institutional Review Board Statement: The study was approved by the Ethics Committee of the Faculty of Arts and Philosophy at Ghent University (30 May 2015).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Abedi, J. Utilizing Accommodations in Assessment. In *Language Testing and Assessment*, *Encyclopedia of Language and Education*; Shohamy, E., Ed.; Springer: Boston, MA, USA, 2017; pp. 2462–2478.
2. Menken, K. NCLB and English Language Learners: Challenges and Consequences. *Theory Pract.* **2010**, *49*, 121–128. [CrossRef]
3. Shohamy, E. Assessing Multilingual Competencies: Adopting Construct Valid Assessment Policies. *Mod. Lang. J.* **2011**, *95*, 418–429. [CrossRef]
4. Abedi, J.; Courtney, M.; Leon, S. *Effectiveness and Validity of Accommodation for English Language Learners in Large-Scale Assessments*; Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California: Los Angeles, CA, USA, 2003.
5. Hofstetter, C.H. Contextual and mathematics accommodation test effects for English-language learners. *Appl. Meas. Educ.* **2003**, *16*, 159–188. [CrossRef]
6. Council of Europe. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*; Language Policy Division; Cambridge University Press: Cambridge, UK, 2018.
7. Butler, F.A.; Stevens, R. *Accommodation Strategies for English Language Learners on Large-Scale Assessments: Student Characteristics and Other Considerations*; Center for Research on Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California: Los Angeles, CA, USA, 1997; Volume 448.
8. Educational Testing Service. *Guidelines for the Assessment of English Language Learners*; Educational Testing Service: Princeton, NJ, USA, 2009.
9. Castellon-Wellington, M. *The Impact of Preference for Accommodations: The Performance of English Language Learners on Large-Scale Academic Achievement Tests*; CSE: Colombo, Sri Lanka, 2000.
10. Rivera, C.; Collum, E.; Shafer Willner, L.; Sia, J.K., Jr. An analysis or state assessment policies in addressing the accommodation of English language learners. In *State Assessment Policy and Practice for English Language Learners: A National Perspective*; Rivera, C., Collum, E., Eds.; Lawrence Erlbaum: Mahwah, NJ, USA, 2006; pp. 1–173.
11. Kieffer, M.J.; Lesaux, N.K.; Rivera, M.; Francis, D.J. Accommodations for English Language Learners Taking Large-Scale Assessments: A Meta-Analysis on Effectiveness and Validity. *Rev. Educ. Res.* **2009**, *79*, 1168–1201. [CrossRef]
12. Pennock-Roman, M.; Rivera, C. Mean effects of test accommodations for ELLs and Non-ELLs: A meta-analysis of experimental studies. *Educ. Meas. Issues Pract.* **2011**, *30*, 10–28. [CrossRef]
13. Kopriva, R.J.; Emick, J.E.; Hipolito-Delgado, C.P.; Cameron, C.A. Do proper accommodation assignments make a difference? Examining the impact of improved decision making on scores for English Language Learners. *Educ. Meas. Issues Pract.* **2007**, *26*, 11–20. [CrossRef]
14. Reed, D.K.; Swanson, E.; Petscher, Y.; Vaughn, S. The effects of teacher read-alouds and student silent reading on predominantly bilingual high school seniors' learning and retention of social studies content. *Read. Writ.* **2013**, *27*, 1119–1140. [CrossRef] [PubMed]
15. Buzick, H.; Stone, E. A Meta-Analysis of Research on the Read Aloud Accommodation. *Educ. Meas.-Issues Pract.* **2014**, *33*, 17–30. [CrossRef]
16. Acosta, B.D.; Rivera, C.; Shafer Willner, L. *Best Practices in State Assessment Policies for Accommodating English Language Learners: A Delphi Study*; The George Washington University Center for Equity and Excellence in Education: Arlington, VA, USA, 2008.

17. Mueller Gathercole, V.C.; Thomas, E.M.; Jones, L.; Guasch, N.V.; Young, N.; Hughes, E.K. Cognitive effects of bilingualism: Digging deeper for the contributions of language domincance, linguistic knowledge, socio-economic status and cognitive abilities. *Int. J. Biling. Educ. Biling.* **2010**, *13*, 617–664. [CrossRef]
18. Gielen, S.; Bellens, K.; Belfi, B.; Van Damme, J. *Het Vierde Leerjaar Basisonderwijs in Vlaanderen: Resultaten van TIMSS 2011 in Internationaal Perspectief En in Vergelijking Met TIMSS 2003*; Centre for Educational Effectiveness and Evaluation: Leuven, Belgium, 2012.
19. OECD. *PISA 2015 Results (Volume I): Excellence and Equity in Education*; OECD Publishing: Paris, France, 2016.
20. Van Laere, E.; Aesaert, K.; Van Braak, J. The role of students' home language in science achievement: A multilevel approach. *Int. J. Sci. Educ.* **2014**, *36*, 2772–2794. [CrossRef]
21. OECD. *Immigrant Background, Student Performance and Students' Attitudes towards Science*; OECD Publishing: Paris, France, 2016.
22. OECD. *PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance to Succeed (Volume II)*; OECD Publishing: Paris, France, 2013.
23. Stansfield, C.W. Oral translation as a test accommodation for ELLs. *Lang. Test.* **2011**, *28*, 401–416. [CrossRef]
24. Erikson, R.; Goldthorpe, J.H.; Portocarero, L. Intergenerational Class Mobility in Three Western European Societies: England, France andSweden. *Br. J. Sociol.* **1979**, *30*, 415–441. [CrossRef]
25. De Houwer, A. Parental language input patterns and children's bilingual use. *Appl. Psycholinguist.* **2007**, *28*, 411–424. [CrossRef]
26. Duursma, E.; Romero-Contreras, S.; Szuber, A.; Proctor, P.; Snow, C.; August, D.; Calderón, M. The role of home literacy and language environment on bilinguals' English and Spanish vocabulary development. *Appl. Psycholinguist.* **2007**, *28*, 171–190. [CrossRef]
27. De Backer, F. *MARS. Meertaligheid Als Realiteit op School. Eindrapport*. 2015.
28. Heppt, B.; Haag, N.; Böhme, K.; Stanat, P. The Role of Academic-Language Features for Reading Comprehension of Language-Minority Students and Students from Low-SES Families. *Read. Res. Q.* **2015**, *50*, 61–82. [CrossRef]
29. Abedi, J.; Ewers, N. *Accommodations for English Language Learners and Students with Disabilities: A Research-Based Decision Algorithm*; Smarter Balanced Assessment Consortium: Santa Clara, CA, USA, 2013.
30. Farrell Hunter, M.; Davidson, M.; Osenga, T. The Simple View of Reading | Reading Rockets. Available online: https://www.readingrockets.org/topics/about-reading/articles/simple-view-reading (accessed on 11 December 2023).
31. Cohen, D.; Tracy, R.; Cohen, J. On the Effectiveness of Pop-Up English Language Glossary Accommodations for EL Students in Large-Scale Assessments. *Appl. Meas. Educ.* **2017**, *30*, 259–272. [CrossRef]
32. Jimerson, S.R.; Ferguson, P. A longitudinal study of grade retention: Academic and behavioral outcomes of retained students through adolescence. *Sch. Psychol. Q.* **2007**, *22*, 314–339. [CrossRef]
33. OECD. *Education at a Glance 2016*; OECD Publishing: Paris, France, 2016.
34. Prevoo, M.J.L.; Malda, M.; Mesman, J.; van IJzendoorn, M.H. Within- and Cross-Language Relations between Oral Language Proficiency and School Outcomes in Bilingual Children with an Immigrant Background: A Meta-Analytical Study. *Rev. Educ. Res.* **2016**, *86*, 237–276. [CrossRef]
35. O'Reilly, T.; McNamara, D.S. The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. *Am. Educ. Res. J.* **2007**, *44*, 161–196. [CrossRef]
36. Yore, L.; Bisanz, G.L.; Hand, B.M. Examining the literacy component of science literacy: 25 years of language arts and science research. *Int. J. Sci. Educ.* **2003**, *25*, 689–725. [CrossRef]
37. Cummins, J. BICS and CALP: Empirical and Theoretical Status of the Distinction. In *Literacies and Language Education*; Springer: Cham, Switzerland, 2017; pp. 59–71.
38. Kuncel, N.R.; Credé, M.; Thomas, L.L. The Validity of Self-Reported Grade Point Averages, Class Ranks, and Test Scores: A Meta-Analysis and Review of the Literature. *Rev. Educ. Res.* **2005**, *75*, 63–82. [CrossRef]
39. Rosen, J.A.; Porter, S.R.; Rogers, J. Understanding Student Self-Reports of Academic Performance and Course-Taking Behavior. *AERA Open* **2017**, *3*, 2332858417711427. [CrossRef]