Sophie Martini, Christine Schiltz, Antoine Fischbach, and Sonja Ugen

Identifying math and reading difficulties of multilingual children: Effects of different cut-offs and reference groups

1 Introduction

An increasing number of students speak a language at home that differs from the language of instruction at school (L2 students). It is estimated that roughly half of the children in the world learn to read in a language other than their home language and are taught in an L2 (McBride-Chang, 2004). In this chapter, we use L2 to refer to all other languages children learn or speak that are different from the main language of instruction. Some children, regardless of which language they speak at home, encounter severe problems with reading or math and may have a specific learning disorder (SLD). Identifying students with SLD can be challenging, as it often co-occurs with other disorders (Fletcher et al., 2018). Identifying L2 students with SLD can be even more challenging, because these students' L2 proficiency often develops differently from students who do speak the language of instruction at home (L1 students), and weak L2 language proficiency has to be ruled out as a cause of low achievement on diagnostic tests administered in an L2 (American Psychiatric Association, 2013). Although, diagnostic criteria and normed diagnostic tests for SLD exist, their norms are often based on an overrepresentation of L1 students. Therefore, these norms may not be accurate for L2 students due to their different language development from L1 students. In this chapter we will explain what SLD is and why it is especially challenging to identify L2 students with SLD.

Early diagnosis and identification of SLD are paramount and screeners may help determine which students to refer for further diagnostic testing. In this chapter, we do not aim to identify and label students with a clinical diagnosis, such as SLD as specified in the *DSM-5* (American Psychiatric Association, 2013). Rather, we aim to identify students who are in the lowest achievement groups of large-scale tests and may need to be referred for further diagnostic SLD testing. We will use the terms *learning difficulties* (LDs), *reading difficulties* (RD),

and *math difficulties* (MD) to refer to these groups. Concretely, we explore what would happen if a large-scale math and reading test would be used as a screener for children at risk for MD and RD in a multilingual education setting. We investigate how different cut-off settings, that is, cut-offs at different percentiles and with different language reference groups, impact the profile of students characterized as having LD. Lastly, we discuss the implications of our findings in relation to (diagnostic) testing in general.

1.1 What is SLD?

SLD is listed in the *DSM-5* as a neurodevelopmental disorder (American Psychiatric Association, 2013). There are different subtypes of SLD, namely, with impairment in reading, written expression, or in math. Each subtype can manifest itself in different ways. For example, SLD with impairment in reading may correspond to a diagnosis of impairments in word reading accuracy, reading rate or fluency, and/or reading comprehension. For impairments in math, students may have problems with number sense, memorization of arithmetic facts, accurate or fluent calculation, and/or correct math reasoning. To start a diagnostic process, students' achievement should be substantially lower than expected for their age and this lower achievement should persist for at least six months. An SLD diagnosis can be given only after a child has been tested individually, using appropriate standardized tests. Furthermore, an SLD should not be explained by a low proficiency in the language of instruction (American Psychiatric Association, 2013: 66–67). Impairments in math and in reading are assumed to have equal prevalence rates (e.g., Geary, 1993) of less than 10% (Desoete et al., 2004; Gross-Tsur et al., 1996).

1.2 Identification of SLD

Identifying students with SLD may be complicated in general, but it is even more so for those who do not speak the language of instruction at home while being tested in the language of instruction, as low proficiency in the language of instruction has to be excluded as a potential cause. Students are usually tested in their language of instruction, as that is the language in which they learned how to read/write and calculate. Due to L2 students' different language development compared to L1 students, they may have lower language of instruction proficiency than their L1 peers. Diagnostic tests are rarely normed with L2 students as a separate reference group, which can lead to both over-identification (e.g. Cummins, 1984, cited

in Limbos & Geva, 2001) and under-identification (e.g. Limbos & Geva, 2001) of L2 students. On the one hand, L2 students may thus have lower proficiency in the language of instruction than L1 students, and this lower proficiency may be classified as an SLD according to the test norms. On the other hand, if it is assumed that students' difficulties are caused by a low L2 proficiency, difficulties caused by a possible SLD may be missed. The interaction between language of instruction proficiency and a possible SLD is especially important to consider when diagnosing reading disorders. However, math learning is also highly influenced by language, for example, counting and transcoding (Kempert et al., 2019). Hence, language proficiency also impacts the diagnosis of learning disorder in mathematics.

It is assumed that the prevalence of SLD is the same for L2 students as for L1 students (Letts, 2011), yet on large-scale achievement tests L2 students often lag behind their L1 peers. Furthermore, L2 students are often referred to as a homogeneous group, though that may not be accurate and may mask differential characteristics between L2 students in terms of home language types and SES (Jang et al., 2013). Considering differences in home languages and SES is important, because they underlie reading comprehension in an L2 (Geva & Wiener, 2014). Socioeconomic status (SES) is a proxy for the resources students have at their disposal and can comprise, for example, parental education, income, or possessions at home (Lenkeit et al., 2018). SES may be related to math and reading development, as SES affects students' language development (Hammer et al., 2014; Hoff, 2006, 2013), their numerical abilities (Mejias & Schiltz, 2013), and academic achievement (e.g., Pace et al., 2017; Paetsch et al., 2015).

1.3 Cut-offs to screen for SLD

A wide variety of cut-offs are used to screen for SLD. Screening tests are often group-based and identify students with low performance. It is quite common to label students in the lowest achievement group of a large-scale test as (potentially) being at risk for developing SLD. A certain percentile (cut-off) on the frequency distribution of the test scores is defined and students at or below this percentile are flagged as at risk for developing SLD. For instance a cut-off at the 10th percentile means that 10% of the test takers will be labeled at risk for developing SLD or as having learning difficulties. The test score corresponding to the 10th percentile is the cut-off score. However, there is no consensus on which cut-off should be used to classify this lowest-achieving group with difficulties, especially in multilingual populations. In reading studies, the 25th percentile is often used, whereas the 10th percentile is very common in math research. It is generally accepted that cut-offs above the 25th percentile are undesirable, as too many

students with average ability would be flagged (Fletcher et al., 2018). Therefore, when examining MD or RD, these two cut-offs are the most obvious ones. Although, the 10th percentile is probably most suitable to screen children for SLD as it appears to be more stable across time (Braeuning et al., 2020). Additionally, less than 10% of the population is estimated to have SLD (Desoete et al., 2004; Gross-Tsur et al., 1996). However, the 25th percentile may still be very useful to find the group of children who need extra support, but who may not have SLD.

Besides which cut-offs are most suitable, there is also no consensus on which *reference group* should be used for norming diagnostic tests. Combined, the cut-off and the reference group used are called "cut-off setting" in this chapter. Theoretically, norms should be based on a reference group that is representative of and comparable to the characteristics of the tested student, at least for those characteristics that may influence students' performance (e.g., age, gender, grade, proficiency in the test language) (American Educational Research Association et al., 2014). In the following section, we will present three different reference groups and describe how their use affects the identification of students with SLD.

(1) Sometimes the whole sample that is available is used as a reference group, without differentiating for students' characteristics such as SES or home language (such as for the *Tempo-Test-Rekenen* (TTR); an arithmetic test) (de Vos, 1992). If this sample is representative of the population, this may lead to L1 students with SLD performing above the cut-offs and L2 students without SLD below the cut-off, due to the difference in language of instruction development.

(2) In countries or regions with one main language of instruction, which is the L1 for most students, tests may be normed based on (an overrepresentation of) students for whom the language of instruction is their L1 (e.g., Leysen et al., 2018). This could lead to norms that identify the expected proportion of L1 students with difficulties, but a larger number of L2 students as their test scores are often lower.

(3) Lastly, it has been suggested that students who speak a language other than the language of instruction should be compared to each other, instead of to the whole sample or their L1 peers (Bedore & Pena, 2008). This is done to try to compare students to other students with similar backgrounds, and thus similar development of the language of instruction development, to each other. Ideally, this would lead to the expected and similar proportions of L1 and L2 students below the cut-offs, as the same proportion of students is expected to have SLD, regardless of language background. This way of norming is, for example, implemented for the *Diagnostischer Rechtschreibtest*-tests (e.g. Müller, 2004a, 2004b) that have norms for German home language students, and for students who speak another language than German at home. This norming

does not take into account the specific L2s of the students, though different students' L2s may affect language of instruction development differently (e.g., Geva & Wiener, 2014).

1.4 The present study

The present chapter focuses on the effect of setting criteria for the screening of learning difficulties by using a complete population dataset. We do not focus on clinical SLD, as the tests used in this chapter were not primarily designed to identify SLD; however, they can be considered as potential screeners to identify children who have math or reading difficulties and may be at risk of developing SLD. These group-based tests can only be considered as global achievement indicators of reading and math. They cannot give precise information on the cognitive processes underlying reading or calculation which would be necessary to diagnose SLD. Most screening tests are group-based (e.g., Mejias et al., 2019; Nosworthy et al., 2013) followed by a more precise, diagnostic, individual follow-up.

Most cut-offs on standardized large-scale tests used to identify potential math and reading difficulties are based on samples that consist of students who speak the language of instruction at home. The present study aims to investigate the impact of different cut-offs and reference groups on the number of children being considered as having reading difficulties (RD) and math difficulties (MD) in relation to third graders' linguistic backgrounds. In this chapter we look at RD and MD separately. Due to space limitations, the analyses for combined RD and MD could not be included in this chapter.

More precisely, we investigated the effect of using different cut-offs on L2 students (of different home language groups) and bilingual students (L1 and another language at home) compared to L1 students (who speak the language of instruction at home) in the multilingual educational setting of Luxembourg. In Luxembourg, there are currently no standardized diagnostic tests that are normed on the country's multilingual population, and that take the trilingual educational system into account. In the Luxembourgish school system, kindergarten is taught in Luxembourgish. From grade 1 onward, students acquire formal literacy skills in German, while learning this language; German functions as the main language of instruction at the same time. The Luxembourgish student population is also very multilingual, as the majority of students speak a language other than Luxembourgish or German at home. The main other languages spoken at home are Portuguese, French, and South Slavic languages (Ministry of National Education, Children and Youth, Department of Statistics and Analysis, 2018). Though this level of multilingualism at a national level may be quite exceptional, other

countries (for example Canada) (Martel et al., 2011) are facing an increasingly diverse school population, especially in urban areas (see, for example, the sample from Geva & Yaghoub Zadeh (2006)). We compared different ways of setting cut-offs to identify students with MD and RD. For that reason, we used a dataset that comprises the population and chose the two most common cut-offs (10th and 25th percentiles) and three different reference groups (whole sample, native (L1), and within each home language group).

We aimed to answer the following research question: What is the effect of different cut-off settings on cut-off scores and consequently on the amount and the characteristics of students classified as having MD and RD?

2 Method

2.1 Participants

The data were collected as part of the Luxembourgish National School Monitoring Programme in 2016. The tests in this program are administered to all students enrolled in Luxembourgish state-funded schools that follow the national curriculum. The overall sample included 5367 third-grade students. This is the vast majority of third graders in Luxembourg, as there are few private schools (Ministry of National Education, Children and Youth, Department of Statistics and Analysis, 2018). In this analysis, we excluded students whose math, German reading comprehension (RC), and/or German listening comprehension (not used in this study) scores were missing (N = 169) and/or whose sex indication was missing (N = 87). The final sample contained 5111 students (49.7% girls).

Students were divided into six home language groups (HLGs), based on which language(s) the student reported to speak with their two primary caretakers, for example, their mother and father (see Tab. 1). These HLGs comprise all students in the Luxembourgish public school system that speak these languages at home and are therefore representative. For four HLGs, both parents mainly speak the same language to their child: Luxembourgish, French, Portuguese, or South Slavic. Students in the South Slavic group speak South Slavic languages of former Yugoslavia with their parents, that is, Bosnian, Croatian, Macedonian, Montenegrin, Serbian, and Serbo-Croatian. For two HLGs, parents mainly speak two different languages to their child: Luxembourgish and French, and Luxembourgish and Portuguese. Luxembourgish and German were grouped together for the establishment of the home language groups, as these languages are linguistically very close (Serva & Petroni, 2008) and are considered L1 students in

Tab. 1: Overview of the cut-off settings, the reference groups, and the number of students per cut-off setting. For the six home language groups (HLGs) the mean HISEI scores and number of students whose HISEI score is missing are listed.

Cut-off settings	Reference groups	N	%	Cut-off percentiles		HISEI			
						Mean	Std. dev.	No. of missings	%Missings in HLG
Population	Whole sample	5111	100	10	25				
HLG	Luxembourgish	1568	31	10	25	53	15	167	11
	French	415	8	10	25	54	15	57	14
	Portuguese	1195	23	10	25	34	12	256	21
	South Slavic	219	4	10	25	37	13	55	25
	Luxembourgish-French	259	5	10	25	54	14	26	10
	Luxembourgish-Portuguese	208	4	10	25	43	14	36	17
Native	Luxembourgish HLG	1568		10	25				

this study. Additionally, German-speaking children (with both parents) constitute only 1.3% (N = 67) of the sample. The six HLGs constitute 75.6% of the sample; the remaining 24.4% of students speak various other languages at home and were excluded from the analyses. These students' language backgrounds are too diverse to be grouped together, but too small to be able to calculate reliable cut-offs for each home language background separately.

Students' SES was based on the professions caretakers reported on the background questionnaire. The profession of the caretaker that ranks highest on the international socioeconomic index of occupational status (HISEI) scale (Ganzeboom et al., 1992) was taken as the SES-indicator. An overview of the mean HISEI scores per HLG can be found in Tab. 1.

2.2 Instruments

Competency Measures. The school monitoring program in grade 3 consisted of three paper-pencil-based competency tests: mathematics (divided over two sessions), German reading comprehension, and German listening comprehension

(not reported here). These tests were standardized within cohorts as well as between cohorts of different years. Students' parents also filled out a background questionnaire, on which students' SES scores were based. The competency tests were taken at students' schools. Each test session lasted for 50 min. The items of all competency tests were scaled with a unidimensional Rasch model. The resulting estimates were converted to standardized scores (Lorphelin et al., 2014).

German Reading Comprehension (RC). The German RC test took 50 minutes. In this test, "[closed and half-open] items mainly address two sub-competencies: (a) locating and understanding written information, and (b) interpreting written information and applying reading strategies" (Sonnleitner et al., 2014: 8). For this measure we used the scale as it is used for school monitoring purposes with a reliability of .892.

Math. The math test as used in the school monitoring program comprised "two content domains: (a) numbers and operations, and (b) space and shape. Item development further covers two contexts (applied vs. not applied)" (Sonnleitner et al., 2014: 8). A subset of the math items was used for this study: only items that test students' ability to do calculations were included. The selected math items were as language-free as possible; that is, students could solve them without reading an explanation and did not need to know specific terms in German to solve the tasks. This was done so that students' math achievement is affected by reading comprehension and German-language proficiency as little as possible. Of the 75 items in the math test, 38 items met the criteria for this scale. The final scale contained 35 items, because three items had to be removed due to poor model fit (Fischbach et al., 2014). The reliability of this math scale is .605.

2.3 Groupings

This chapter explores the impact of two different cut-offs (percentiles 10 and 25) for RC and math achievement for three different reference groups. (1) The first reference group is the whole sample; that is, the cut-off scores have been calculated based on the achievement of all students who took part in the test. (2) The second reference group was within the home language groups; that is, the cut-offs were based on the performance within each of the 6 HLGs. (3) The third reference group was the native language group. In an educational system where there is one dominant language that is used as the language of instruction and most students are raised in this language, diagnostic and large-scale tests are often standardized based on a supposedly mostly native, monolingual, majority language group. In this context, the native reference group is the Luxembourgish HLG.

Students who performed below the cut-offs (percentile 10 or 25) were classified as having potential learning difficulties. Due to rounding up or down to whole percentages, and because a number of students may have the same test score, the actual number of students below the cut-off based on the HLGs and the native reference group may be slightly higher or lower than the expected 10 and 25%. An overview of these cut-off settings can be found in Tab. 1.

2.4 Procedure

Data were collected in five testing sessions. The tests were administered by the students' teachers, who had received instruction manuals on how to administer these tests. The national school monitoring program has a proper legal basis and has been approved by the national committee for data protection. Parents were asked to fill out a background questionnaire on, for example, their educational and occupational backgrounds and country of birth. All students and their parents or legal guardians were duly informed before the data collection and had the possibility to opt out. All statistical analyses were performed with anonymized data.

3 Results

In this section we will discuss the differences in prevalence of math and reading difficulties for the six different HLGs. We will first focus on the exact cut-off scores for all six cut-off settings (Section 3.1). In a second step, we will examine the prevalence of MD and RD for the six HLGs (Section 3.2).

3.1 Cut-off scores

Before looking at the cut-offs for MD and RD, and consequently the prevalence of these difficulties for students from the six different HLGs, differences in mean scores between different HLGs are described in Tab. 2. Differences in mean scores are related to cut-off scores and number of students below the cut-offs, depending on the reference group that is chosen. On average, students in the Luxembourgish and Luxembourgish-French HLGs have the highest German RC and math scores. The other four HLGs have lower mean scores.

For the native reference group and the whole sample reference group, at the 10th and 25th percentile cut-offs for both math and German RC, the scores

Tab. 2: The mean scores and cut-off scores per cut-off setting for all reference groups.

Cut-off settings	Reference groups	Mean scores per reference group		Cut-off scores per percentile			
				10		25	
		Math	RC	Math	RC	Math	RC
1) Whole sample	Whole sample	501	501	382	325	433	397
2) HLG	Luxembourgish	522	584	395	397	457	493
	French	504	465	368	325	433	384
	Portuguese	475	415	368	291	421	341
	South Slavic	483	467	354	341	421	384
	Luxembourgish-French	514	541	395	384	445	451
	Luxembourgish- Portuguese	461	446	354	291	408	356
3) Native	Luxembourgish HLG	522	584	395	397	457	493

based on the native reference group are always higher than the whole sample reference group. These differences are bigger for German RC than for math, and for the 25th percentile than for the 10th percentile; see Tab. 2.

For the cut-off scores within the HLGs as reference groups, there is a cut-off score for each HLG. In this case, the proportion of students identified as having difficulties is the same for each HLG at each cut-off, but the cut-off score will differ depending on the distribution of math and RC scores; the cut-off scores differ within each HLG.

Table 2 shows that for the cut-off scores with the HLGs as reference group, the cut-off scores for the Luxembourgish HLG are higher than the cut-off scores for the other home language groups. The only exception is the math 10th percentile cut-off score of the Luxembourgish-French HLG (395), which is equal to the Luxembourgish HLG (395) cut-off. Overall, the differences between the MD cut-offs are smaller than the differences between the RC cut-offs of the six HLGs, and the differences between the 10th percentile cut-offs are smaller than the 25th percentile cut-offs. However, the differences between cut-off scores for the different HLGs can be large: for example 152 points between the Luxembourgish (493) and Portuguese (341) HLGs at the 25th percentile for RD.

For most HLGs, the cut-off score for math is higher than the one for RC at the same percentile. Moreover, some HLGs have the same or similar cut-off scores for one subject, but not for the other. This implies that L2 and language

proficiency interact differently, depending on the HLG. For example, the 10th percentile cut-off scores for math are equal for the French and Portuguese HLGs, namely, 368, but different for RC, namely, 325 and 291. Additionally, for the 25th percentile cut-off scores, the Portuguese and South Slavic HLGs' math cut-off scores is 421, but their RC cut-off scores are 43 points apart.

In sum, the six different cut-off settings result in different cut-off scores. As a consequence of these differences in cut-off scores, the prevalence of math and reading difficulties varies per cut-off setting.

3.2 Prevalence of MD and RD

The prevalence of math and reading difficulties varies per cut-off setting. For the cut-off settings with the whole sample and within HLGs as reference groups, respectively 10 or 25% of the students (either for the whole sample or for the six HLGs) are classified as having difficulties. However, for the cut-off settings based on the native reference group, only the proportion of students in the Luxembourgish HLG is fixed at 10 or 25%, but not for the other five HLGs. Therefore, the percentage of students with MD and RD in the other five HLGs may be higher or lower than 10 and 25 percent respectively.

If the students with difficulties were distributed equally over all HLGs, 25 or 10% respectively of students in all home language groups would be classified as having difficulties for the two cut-off settings based on the whole sample and the two settings based on the Luxembourgish home language group. However, as is shown in Fig. 1 (and Appendix A), this is not the case.

For the cut-offs with the HLGs as reference groups, the percentage of students below the cut-offs is fixed for each HLG at 10 or 25 percent respectively. Therefore, the cut-offs based on the HLGs as reference group are not discussed in this section. However, the cut-off scores differ between HLGs, which is discussed in Section 3.1.

3.2.1 Whole sample reference group

25th percentile
For MD, the Luxembourgish and Luxembourgish-French HLGs have fewer students with MD than the expected 25%: 19 and 21; see Fig. 1, a2. The French home language group would have 27% of students with MD, slightly more than expected. The Portuguese, South Slavic, and Luxembourgish-Portuguese HLGs have the most students that are identified with MD: 34, 36, and 40% respectively. For RD

(Fig. 1, a1), the difference in proportion of students per HLG that are classified as having difficulties are larger than for MD. Again, the Luxembourgish and Luxembourgish-French HLGs have fewer students with RD than the expected 25%: only 10 and 13%. The French and South Slavic HLGs have 30 and 31% of students with RD, which is slightly more than the expected 25%. However, the Portuguese and Luxembourgish-Portuguese HLGs have more students with RD than expected, namely, 48 and 37%.

10th percentile

For MD, the Luxembourgish and Luxembourgish-French HLGs have fewer students with MD than the expected 10%: 7 and 6%, as shown in Fig. 1, a2. The French HLG would have 11% of students with MD, slightly more than expected. The Portuguese, South Slavic, and Luxembourgish-Portuguese HLGs have the most students that are identified with MD: 13, 13, and 14% respectively. This pattern is in line with the one found for the 25th percentile whole sample cut-off. For RD, there is a bigger difference in the proportion of students classified as having difficulties than for MD.
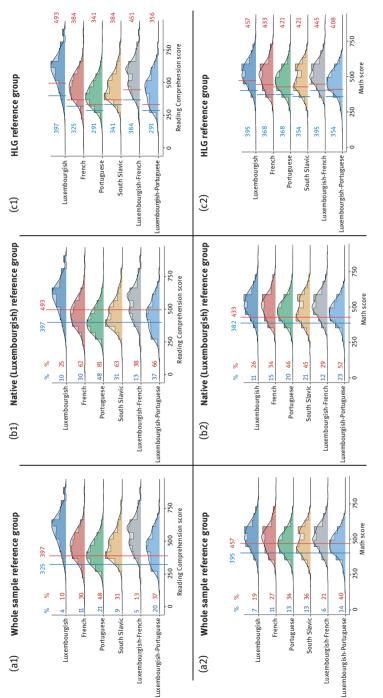
Again, the Luxembourgish and Luxembourgish-French HLGs have the smallest proportion of students with RD (Fig. 1, a1), namely, 4 and 5%. The French HLG has 11% students with RD, slightly more than the expected 10%; the South Slavic HLG has 9% students with RD, slightly below the expected 10%. Similar to MD, the HLGs with the highest proportion of students with difficulties are the Portuguese and Luxembourgish-Portuguese HLGs: 21 and 20%.

3.2.2 Native reference group

25th percentile

As this cut-off setting is based on the Luxembourgish HLG, this group has 26% (not 25, due to duplicate scores and rounding up to a whole percentage) of students with MD. The other HLGs have a higher proportion of students with MD. The French and Luxembourgish-French HLGs are closest to the expected 25% of students with MD, namely, 34 and 29%; see Fig. 1, b2. Similar to the 25th percentile cut-off with the whole sample reference group, the Portuguese, South Slavic, and Luxembourgish-Portuguese HLGs have the highest proportions of students with MD: 46, 45, and 52%.

For RD, all other HLGs have more students with RD than the reference group. The Luxembourgish-French HLG has the lowest proportion of students with RD after the Luxembourgish HLG: 38%. Three groups have a similar proportion of

Fig. 1: Cut-off scores for reading comprehension (a1, b1, c1) and math scores (a2, b2, c2) for the whole sample (a1, a2), native (Luxembourgish) (b1, b2), and HLG reference groups (c1, c2). Per reference group, the cut-off score at the 10th percentile is shown in blue and the cut-off score at the 25th percentile in red. For the whole and native language reference groups, percentages of students below the 10th and the 25th percentiles are indicated in blue and red.

students with RD, the French, South Slavic, and Luxembourgish-Portuguese groups: 62, 63, and 66%. The group with most students with RD for this cut-off setting is the Portuguese home language group with 81%.

10th percentile

As this setting is based on the Luxembourgish HLG, 11% of the students in the Luxembourgish HLG are classified as having MD. The Luxembourgish-French HLG have a slightly higher proportion of students with MD, namely, 12%. The group with the next lowest proportion of MD are the French HLG with 15%. The other three HLGs, that is, the Portuguese, South Slavic, and the Luxembourgish-Portuguese, have the highest proportion of students with MD: 20, 21, and 23%. For RD, again the Luxembourgish HLG has the lowest proportion of students with MD, 10%, as expected for this cut-off. The Luxembourgish-French HLG has a similar proportion of students with RD, 13%. The French, South Slavic, and Luxembourgish-Portuguese HLGs have similar prevalence of RD: that is, 30, 31, and 37%. The HLG with most RD is the Portuguese HLG, namely, 48%.

In sum, for the four cut-off settings discussed above (whole sample reference group, both 10th and 25th percentiles, and native reference group, both 10th and 25th percentiles), the Luxembourgish HLG has the smallest proportion of students with both MD and RD; in all other HLGs, MD and RD are more prevalent. Sometimes, difficulties are only slightly more frequent than the Luxembourgish HLG than in other HLGs, for example, MD for the Luxembourgish-French and French HLGs, for most of the four cut-off settings discussed. However, for one of the other HLGs, the proportion of students classified with MD is twice as high as in the Luxembourgish HLG, namely, for the Portuguese HLG. Though the pattern of HLGs with more or less MD and RD is similar for these four cut-offs, the actual proportion of students classified as having difficulties varies, as the exact cut-off score differs (see Tab. 2)

3.3 Consistency of identified students across cut-off settings

It is interesting to examine how many students are classified as having difficulties when using all three reference groups, versus how many are only classified as such for one or two reference groups. This will provide further insights into how the different cut-off settings work for the different HLGs.

3.3.1 RD prevalence across the cut-offs

Across the different cut-off settings, a different pattern emerges for the six HLGs (see Tab. 3). At the 10th percentile, about 10% of students in the French, Portuguese, South Slavic, and Luxembourgish-Portuguese HLGs are consistently below the cut-off for all three reference groups, at the 25th percentile, that is, 25%. The Luxembourgish and Luxembourgish-French HLGs have a lower proportion of students who are below all three cut-offs than the other HLGs. Additionally, these two HLGs are the only ones with students who are only below the HLG and below the native reference group cut-offs. In the French, Portuguese, South Slavic, and Luxembourgish-Portuguese HLGs, there are students below the native and below the whole sample cut-offs; this is not the case for the Luxembourgish and Luxembourgish-French HLGs. Additionally, for these four HLGs many students are only below the native reference group cut-offs: 19, 27, 22, and 17% of students at the 10th percentile and 32, 33, 32, and 29% at the 25th percentile.

3.3.2 MD prevalence across the cut-offs

Similar to RD, a different pattern emerges for the six HLGs across the different cut-off settings. At the 10th and 25th percentiles, about 10 or 25% of students in the French, Portuguese, South Slavic, and Luxembourgish-Portuguese HLGs are below the cut-off for all three reference groups; see Tab. 4. The Luxembourgish and Luxembourgish-French HLGs have a lower proportion of students who are below all three cut-offs. As for RD, these two HLGs are the only groups with students below the HLG and the native reference group cut-offs, but above the whole sample cut-offs. In the French, Portuguese, South Slavic, and Luxembourgish-Portuguese HLGs, there are students below the native and whole sample cut-offs but above the within HLG reference group cut-off; this is not the case for the Luxembourgish and Luxembourgish-French HLG. Additionally, for these four HLGs many students are only below the native reference group cut-offs: 4, 7, 8, and 9% of students at the 10th percentile and 7, 12, 10, and 12% at the 25th percentile.

4 Discussion

In this study we aimed to answer the question: What is the effect of different cut-off settings on the classification of students with MD and RD in a multilingual student population? Concretely, we examined the impact of different language-group

Tab. 3: Percentage (%) of students below one, two, or three reference groups (RG) cut-offs per home language group (HLG) for reading.

		Luxembourgish	French	Portuguese	South Slavic	Luxembourgish-French	Luxembourgish-Portuguese	Overall
10th percentile								
Below three RG cut-offs	%	3.8	9.2	11.0	8.7	4.6	10.6	7.3
Below HLG and below native RG cut-offs	%	6.3	0.0	0.0	0.5	4.2	0.0	2.9
Below whole sample and below native RG cut-offs	%	0.0	1.7	9.9	0.0	0.0	9.1	3.7
Below native RG cut-off only	%	0.0	19.0	27.2	21.9	4.2	16.8	12.9
Not below any cut-off	%	89.9	70.1	51.9	68.9	86.9	63.5	73.2
25th percentile								
Below three RG cut-offs	%	10.1	25.5	26.4	27.9	13.1	27.4	19.0
Below HLG and below native RG cut-offs	%	14.8	0.0	0.0	0.0	12.7	0.0	6.9
Below whole sample and below native RG cut-offs	%	0.0	4.3	21.7	3.2	0.0	9.1	7.8
Below native RG cut-off only	%	0.0	32.0	32.5	32.0	12.4	29.3	17.7
Not below any cut-off	%	75.1	38.1	19.4	37.0	61.8	34.1	48.6

Tab. 4: Percentage (%) of students below one, two, or three reference groups (RG) cut-offs per home language group (HLG) for math.

		Luxembourgish	French	Portuguese	South Slavic	Luxembourgish-French	Luxembourgish-Portuguese	Overall
10th percentile								
Below three RG cut-offs	%	6.6	10.8	9.6	10.5	6.2	9.1	8.3
Below HLG and below native RG cut-offs	%	4.5	0.0	0.0	0.0	3.1	0.0	2.0
Below whole sample and below native RG cut-offs	%	0.0	0.0	2.8	2.3	0.0	4.8	1.3
Below native RG cut-off only	%	0.0	4.3	7.1	8.2	3.1	8.7	3.8
Not below any cut-off	%	88.9	84.8	80.4	79.0	87.6	77.4	84.6
25th percentile								
Below three RG cut-offs	%	19.4	23.1	23.8	23.7	20.5	23.1	21.7
Below HLG and native RG cut-offs	%	7.0	0.0	0.0	0.0	4.6	0.0	3.1
Below whole sample and below native RG cut-offs	%	0.0	3.6	10.5	11.9	0.0	16.8	5.2
Below native RG cut-off only	%	0.0	7.0	11.5	9.6	3.9	12.0	5.8
Not below any cut-off	%	73.7	66.3	54.1	54.8	71.0	48.1	64.2

related criteria of cut-off scores and consequently the number of students below the cut-offs and thus classified as having MD or RD. The associated number and characteristics of students classified as potentially having MD and RD. We used three different cut-off settings, that is, whole sample, HLG, and native reference group, and found that they resulted in different cut-off scores and hence in large variations in the number of students classified as having potential MD and RD.

Generally, the cut-off scores based on the performance of the native reference group were higher than those based on the whole sample. This means that the performance-level below which a student is considered as having MD or RD is higher when natives are used as the reference group compared to the whole sample. As a consequence, more students in general would be identified as having MD and RD with the native group as reference for the cut-off. Within the whole sample and the native reference group, there were differences between the cut-off scores for math and reading comprehension. For the whole sample, the cut-off scores for math were higher than those for reading, whereas for the native reference group the reading cut-off scores were higher than for math. This indicates that the Luxembourgish native reference group had an advantage in the more language-dependent task, that is, reading comprehension, as it was the only group who took the tests in their L1. The comparisons between the cut-off scores within each of the six HLGs showed that there were large differences between cut-off scores of HLGs. Furthermore, for most HLGs the math cut-off score was higher than the RC cut-off score, except for the Luxembourgish HLG, meaning that generally performance-levels under which students were considered as having RD were relatively lower than those for MD. This may be due to the math scale used in this study, as only items with very low language requirements, which did not require students to read and understand a word problem, were included.

As the cut-off scores differed per cut-off setting, the prevalence rate of MD and RD consequently varied between the different groups per cut-off setting too. Moreover, there were large differences in MD and RD prevalence rate between HLGs for the cut-offs based on the whole sample and native reference groups. For these reference groups' cut-offs, the difference between the Luxembourgish and Portuguese HLGs was most pronounced: MD was twice as prevalent in the Portuguese HLG, and RD five times as frequent as in the Luxembourgish HLG.

4.1 Over- and under-identification

In the literature, both over- and under-identification of L2 students with RD and MD have been reported (e.g. Cummins, 1984 cited in Limbos & Geva, 2001). In this study, when the population and native reference groups are

used to set cut-offs, more students from the French, Portuguese, and South Slavic HLGs than the expected 10 and 25% are classified as having difficulties, especially for German reading. On the other hand, when the population reference group is used, it is very likely that students with an LD would remain unidentified within the (native) Luxembourgish HLG.

When the HLGs are used as reference groups the proportion of students with difficulties is the same across HLGs, but the actual cut-off scores vastly differ, up to 150 points, which equals 1.5 standard deviation. Overall, only a minority of students is consistently below the 10th or 25th percentile respectively for math or reading for all three reference groups. For the Luxembourgish HLG, the students who are not always classified as having difficulties perform above the population cut-off, but below the native/HLG reference group cut-off scores (that is the same for this HLG). In contrast, for the Portuguese HLG, for example, students who are only sometimes classified as having difficulties are above the HLG cut-off, but below the native or population reference group cut-off scores (or both). Therefore, a combination of cut-off settings while considering students' linguistic backgrounds may be most effective to identify students with LDs. Knowing a student's score in relation to the whole sample or native reference group, as well as in comparison to a group of students with a similar background, may help in deciding what follow-up testing or interventions should be taken.

Two HLGs speak two languages at home, of which one is Luxembourgish, yet their cut-off scores and prevalence of MD and RD differ substantially from each other. For the Luxembourgish-French HLG, the cut-off scores within HLG are close to the Luxembourgish HLG, as well as the proportion and ratio of MD and RD. Both these HLGs have relatively high mean SES. In contrast, the Luxembourgish-Portuguese HLG is more like the Portuguese HLG, in terms of mean SES. It also resembles the Portuguese HLG's cut-off scores based within the HLG reference group, and prevalence and ratio of MD and RD. The Luxembourgish-French HLG thus seems to benefit from speaking Luxembourgish at home, while this is not the case for the Luxembourgish-Portuguese group, who have similar cut-off scores and MD and RD prevalence to the Portuguese HLG. This might be indirectly related to SES.

This chapter did not investigate the influence of SES in combination with HLG on math and reading achievement in detail. Previous research with grade-three data from the Luxembourgish National School Monitoring Programme showed that the differences in math and German RC achievement are for a large part explained by home language and SES: students who speak Luxembourgish or German at home and who have high SES have higher mean achievement for both RC and math compared to students with average or low SES and who do not speak either of these languages at home. National School Monitoring data further have

shown that the effect of home language background is bigger for RC than for math and the effect of SES is similar for math and RC (Muller et al., 2015). Furthermore, school monitoring data has shown that that students' German RC in grade 3 was predictive for their German RC in grade 9 (Sonnleitner et al., 2018). However, the predictive values differed for students with high and low SES and between different HLGs. Students with high SES (top quartile) and students who speak Luxembourgish at home were more likely to have high RC and math in later grades. This is in line with the findings of the present study and implies that the students who do not speak Luxembourgish at home and/or come from HLGs with lower mean SES are more likely to fall below cut-offs and have low math and/or reading achievement.

4.2 Screening for difficulties

Screening procedures should identify students that need further diagnostic testing (Fletcher et al., 2018, p. 63). Helping students with a deficit as early as possible usually has the best outcome (Heckman, 2008). For these tests, missing an at-risk child is a bigger problem than having a false positive (i.e., identifying a child that does not have difficulties). Generally, it is indeed less harmful to identify more children than necessary, than missing children who would need further diagnostic testing and help. On the other hand, having too many "false positives" is counterproductive, as valuable resources would be spent on students who may not need it.

Which cut-offs (or a combination of cut-offs) are useful depends on the purpose of the screener: to only select the students who are likely to have a clinical SLD, then the 10th percentile is most suitable. Moreover, at the 10th percentile for the whole sample and HLG reference groups, the prevalence of MD and RD is similar, which is what is expected for clinical SLD in math and reading (e.g. Geary, 1993). However, all students below the 25th percentile might benefit from extra instruction, even though they might not have a clinical SLD. When the whole sample or the native reference groups are used, this may lead to many "false positive" students for SLD; therefore, a comparison with the within HLG cut-off score would be useful to decide what kind of support students need.

In this chapter, we did not discuss the prevalence of MD and RD simultaneously occurring in the same students, due to space limitations. This would be important to consider, as the underlying causes of the learning difficulties may be different in isolated or combined LDs. Therefore, when choosing follow-up

tests for a clinical SLD diagnosis or non-clinical LD identification and the following choice of adequate interventions, this should also be considered.

4.3 Limitations and implications

While a large-scale group-based standardized test can serve as a screener for students at risk of having SLD, these tests are neither designed for nor sufficiently fine-grained to detect clinical SLD. Nevertheless, the results of this study using population data to examine the impact of cut-off scores and reference groups to screen for SLD also have an impact on other settings in which tests are used, such as for diagnostic tests. For instance, the choice of the reference group to establish norms and/or the language of the instruction should be considered to make the diagnostic process more accurate and fair, especially for L2 children. For the SLD diagnosis, several individual follow-up testing sessions with a psychological diagnostic specialist are necessary. For RD, for instance, a potential follow-up could be to investigate reading skills at the word-level, as RD can comprise word decoding or recognition and/or spelling deficiency (Fletcher et al., 2018). Typically L2 students do not differ from L1 students on the level of decoding (Limbos & Geva, 2001), so this type of supplementary assessment could help distinguish between a predominantly German proficiency problem and a clinical RD (e.g., Lesaux & Siegel, 2003).

For the purpose of this study on LDs, we refined the math scale in this study, which artificially decreased the large achievement differences between HLGs usually found (e.g. Muller et al., 2015; Sonnleitner et al., 2018). The adjustment of the math scale was to minimize the language required to solve the items on the math test. This minimization of language explains why the differences between students from six different HLGs were smaller for math achievement than for RC, whereas in previous studies with National School Monitoring data and PISA data these differences are larger. Our findings thus imply that using a subset of language-reduced standardized, large-scale test items may be more fine-grained as a screener for MD in a multilingual setting and should be researched further.

In countries in which national standardized tests are implemented, these tests could potentially serve as a screener for students who are at risk of falling behind. As both math and RC in the language of instruction depend on language of instruction proficiency, strengthening students' language of instruction skills may be helpful for students and may help decrease performance differences and associated cut-off score differences between HLGs. Ertel et al. (2019) found that for Portuguese HLG students in Luxembourg, differences in German

and Luxembourgish language skills were the only language-related factor that differed significantly between low-achieving students who passed second grade and those who were retained. Differences in Portuguese skills were statistically insignificant. In general, providing students who have low RC and/or math achievement in grade 3 with an extended training in the language of instruction could help improve their reading comprehension. Therefore, the large-scale test used in this study or in similar settings could potentially be used to identify students with difficulties, so that they can receive extra help to catch up as soon as possible, to prevent low long-term achievement.

In this study we found that two different cut-offs, that is, the 10th and 25th percentile and three different reference groups (i.e., whole sample, HLGs, native reference group), resulted in different cut-off scores and consequently different proportions of students identified with MD and RD. Practitioners should be aware of all of the caveats of using standardized tests and their corresponding norms when using them.

Appendix A: Math and Reading difficulties per home language group

	25th percentile								10th percentile							
	Whole sample reference group								Whole sample reference group							
	No MD		MD		No RD		RD		No MD		MD		No RD		RD	
	N	% within HL	N	% within HL	N	% within HL	N	% within HL	N	% within HL	N	% within HL	N	% within HL	N	% within HL
Luxembourgish	1264	81	304	19	1409	90	159	10	1465	93	103	7	1508	96	60	4
French	304	73	111	27	291	70	124	30	370	89	45	11	370	89	45	11
Portuguese	785	66	410	34	620	52	575	48	1046	88	149	13	945	79	250	21
South Slavic	141	64	78	36	151	69	68	31	191	87	28	13	200	91	19	9
Luxembourgish and French	206	80	53	21	225	87	34	13	243	94	16	6	247	95	12	5
Luxembourgish and Portuguese	125	60	83	40	132	64	76	37	179	86	29	14	167	80	41	20

Native reference group

	No MD		MD		No RD		RD	
	N	% within HL	N	% within HL	N	% within HL	N	% within HL
Luxembourgish	1155	74	413	26	1177	75	391	25
French	275	66	140	34	158	38	257	62
Portuguese	647	54	548	46	232	19	963	81
South Slavic	120	55	99	45	81	37	138	63
Luxembourgish and French	184	71	75	29	160	62	99	38
Luxembourgish and Portuguese	100	48	108	52	71	34	137	66

Native reference group

	No MD		MD		No RD		RD	
	N	% within HL	N	% within HL	N	% within HL	N	% within HL
Luxembourgish	1394	89	174	11	1409	90	159	10
French	352	85	63	15	291	70	124	30
Portuguese	961	80	234	20	620	52	575	48
South Slavic	173	79	46	21	151	69	68	31
Luxembourgish and French	227	88	32	12	225	87	34	13
Luxembourgish and Portuguese	161	77	47	23	132	64	76	37

(continued)

(continued)

	HLG reference group								HLG reference group							
	No MD		MD		No RD		RD		No MD		MD		No RD		RD	
	N	% within HL	N	% within HL	N	% within HL	N	% within HL	N	% within HL	N	% within HL	N	% within HL	N	% within HL
Luxembourgish	1155	74	413	26	1177	75	391	25	1394	89	174	11	1409	90	159	10
French	319	77	96	23	309	75	106	26	370	89	45	11	377	91	38	9
Portuguese	910	76	285	24	879	74	316	26	1080	90	115	10	1063	89	132	11
South Slavic	167	76	52	24	158	72	61	28	196	90	23	11	199	91	20	9
Luxembourgish and French	194	75	65	25	192	74	67	26	235	91	24	9	236	91	23	9
Luxembourgish and Portuguese	160	77	48	23	151	73	57	27	189	91	19	9	186	89	22	11

Appendix B – List of abbreviations

HLG	Home language group
L1	First language
L2	Second language
LDs	Learning difficulties
MD	Math difficulties
RC	Reading comprehension
RD	Reading difficulties
SLD	Specific learning disorder

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education & Joint Committee on Standards for Educational and Psychological Testing (U.S) (2014): *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

American Psychiatric Association (2013): *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Association. Washington, DC & London, England. https://economie.gouv.cg/sites/default/files/webform/pdf-diagnostic-and-statistical-manual-of-mental-disorders-5th-editio-american-psychiatric-association-pdf-download-free-book-f6714d9.pdf

Bedore, Lisa M. & Pena, Elizabeth D. (2008): Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism* 11 (1), 1–29.

Braeuning, David, Lambert, Katherina, Hirsch, Stefa, Schils, Trudie, Borghans, Lex, Nagengast, Benjamin & Moeller, Korbinian. 2020. Abstract: Diagnose von Rechenschwäche bei Grundschulkindern – Evaluation gängiger Kriterien und Vorgehensweisen. In. Potsdam.

Cummins, Jim. (1984): *Bilingualism and Special Education: Issues in Assessment and Pedagogy*. 6, Taylor & Francis Group. Vol. 6. Clevedon: Multilingual Matters, 1984.

Desoete, Annemie, Roeyers, Herbert & De Clercq, Armand (2004): Children with mathematics learning disabilities in Belgium. *Journal of Learning Disabilities* 37 (1), 50–61.

Ertel, Cintia, Alieva, Aigul, Hornung, Caroline & Schiltz, Christine (2019): *The Effect of Grade Retention on Reading Skills of immigrant children in multilingual elementary school. A longitudinal study*. Esch-sur-Alzette.

Fischbach, Antoine, Ugen, Sonja & Martin, Romain (eds.). 2014. ÉpStan technical report. University of Luxembourg.

Fletcher, Jack M., Reid Lyon, G., Fuchs, Lynn S. & Barnes, Marcia A. (2018): *Learning Disabilities, Second Edition: From Identification to Intervention*. New York, NY: Guilford Publications.

Ganzeboom, Harry BG, De Graaf, Paul M. & Treiman, Donald J. (1992): A standard international socio-economic index of occupational status. *Social Science Research* 21 (1), 1–56.

Geary, David C. (1993): Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin* 114 (2), 345–362. doi:https://doi.org/10.1037/0033-2909.114.2.345.

Geva, Esther & Wiener, Judith (2014): *Psychological Assessment of Culturally and Linguistically Diverse Children and Adolescents: A Practitioner's Guide*. New York, NY: Springer Publishing Company.

Geva, Esther & Zadeh, Zohreh Yaghoub (2006): Reading efficiency in native english-speaking and english-as-a-second-language children: The role of oral proficiency and underlying cognitive-linguistic processes. *Scientific Studies of Reading* 10 (1), 31–57. doi:https://doi.org/10.1207/s1532799xssr1001_3.

Gross-Tsur, Varda, Manor, Orly & Shalev, Ruth S. (1996): Developmental dyscalculia: Prevalence and demographic features. *Developmental Medicine & Child Neurology* 38 (1), 25–33.

Hammer, Carol Scheffner, Hoff, Erika, Uchikoshi, Yuuko, Gillanders, Cristina, Castro, Dina & Sandilos, Lia E. (2014): The language and literacy development of young dual language learners: A critical review. *Early Childhood Research Quarterly* 29 (4), 715–733. doi: https://doi.org/10.1016/j.ecresq.2014.05.008.

Heckman, James J. 2008. Schools, skills, and synapses. NBER Working Paper No. 14064. *National Bureau of Economic Research*.

Hoff, Erika (2006): How social contexts support and shape language development. *Developmental Review* 26 (1), 55–88. doi:https://doi.org/10.1016/j.dr.2005.11.002.

Hoff, Erika (2013): Interpreting the early language trajectories of children from Low SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology* 49 (1), 4–14. doi:https://doi.org/10.1037/a0027238.

Jang, Eunice Eunhee, Dunlop, Maggie, Wagner, Maryam, Kim, Youn-Hee & Zhimei, Gu. (2013): Elementary school ELLs' reading skill profiles using cognitive diagnosis modeling: Roles of length of residence and home language environment. *Language Learning* 63 (3), 400–436.

Kempert, Sebastian, Schalk, Lennart & Saalbach, Henrik (2019): Übersichtsartikel: Sprache als Werkzeug des Lernens: Ein Überblick zu den kommunikativen und kognitiven Funktionen der Sprache und deren Bedeutung für den fachlichen Wissenserwerb. *Psychologie in Erziehung und Unterricht* 66 (3), 176–195. doi:https://doi.org/10.2378/PEU2018.art19d.

Lenkeit, Jenny, Schwippert, Knut & Knigge, Michel (2018): Configurations of multiple disparities in reading performance: Longitudinal observations across France, Germany, Sweden and the United Kingdom. *Assessment in Education: Principles, Policy & Practice* 25 (1), 52–86. doi:https://doi.org/10.1080/0969594X.2017.1309352.

Lesaux, Nonie K. & Siegel, Linda S. (2003): The development of reading in children who speak English as a second language. *Developmental Psychology* 39 (6), 1005.

Letts, Carolyn (2011): Communication impairment in a multilingual context. In Ellis, Sue, McCarthy, Elspeth (eds.): *Applied Linguistics and Primary School Teaching*. Cambridge: Cambridge UP. doi:10.1017/CBO9780511921605.026.

Leysen, Heleen, Van den Broek, Wim, Keuning, Jos, Noé, Marjolein & Geudens, Astrid (2018): Vlaamse Normering van de Drie-Minuten-Toets en AVI-toetskaarten van 2009. Thomas More: Antwerpen.

Limbos, Marjolaine M. & Geva, Esther (2001): Accuracy of teacher assessments of second-language students at risk for reading disability. *Journal of Learning Disabilities* 34 (2), 136–151.

Lorphelin, Dalia, Keller, Ulrich, Fischbach, Antoine & Brunner, Martin (2014): Data processing, analyses, and reporting. In Fischbach, Antoine, Ugen, Sonja, Martin, Romain (eds.): *ÉpStan Technical Report*. Luxembourg: University of Luxembourg, LUCET, 12–29.

Martel, Laurent, Malenfant, Erin Caron, Morency, Jean-Dominique, Lebel, André, Bélanger, Alain & Bastien, Nicolas (2011): Projected trends to 2031 for the Canadian labour force. *Canadian Economic Observer* 24, 6.

McBride-Chang, Catherine (2004): *Children's Literacy Development*. London, United Kingdom: Taylor & Francis Group. http://ebookcentral.proquest.com/lib/unilu-ebooks/detail.action?docID=746361 (8 July, 2020)

Mejias, Sandrine, Muller, Claire & Schiltz, Christine (2019): Assessing mathematical school readiness. *Frontiers in Psychology* 10, 1173. doi:https://doi.org/10.3389/fpsyg.2019.01173.

Mejias, Sandrine & Schiltz, Christine. 2013. Estimation abilities of large numerosities in Kindergartners. 4. 1–12. https://doi.org/10.3389/fpsyg.2013.00518.

Ministry of National Education, Children and Youth, Department of Statistics and Analysis. 2018. The Key Figures of the National Education – Statistics and Indicators 2016/2017.

Muller, Claire, Reichert, Monique, Gamo, Sylvie, Hoffmann, Danielle, Hornung, Caroline, Sonnleitner, Philipp & Martin, Romain (2015). Kompetenzunterschiede aufgrund des Schülerhintergrundes. *Épreuves Standardisées: Bildungsmonitoring für Luxemburg. Nationaler Bericht 2011 bis 2013*, 34–56.

Müller, Rudolf. 2004a. Diagnostischer Rechtschreibtest für 3. *Klassen (DRT-3)*. Göttingen: Hogrefe.

Müller, Rudolf (2004b): *Diagnostischer Rechtschreibtest für 1. Klassen: DRT 1; Manual*. Beltz Test.

Nosworthy, Nadia, Bugden, Stephanie, Archibald, Lisa, Evans, Barrie & Ansari, Daniel (2013): A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. (Ed.) Kevin Paterson. *PLoS ONE* 8 (7), e67918. doi:https://doi.org/10.1371/journal.pone.0067918.

Pace, Amy, Luo, Rufan, Hirsh-Pasek, Kathy & Golinkoff, Roberta Michnick (2017): Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics* 3 (1), 285–308. doi:https://doi.org/10.1146/annurev-linguistics-011516-034226.

Paetsch, Jennifer, Felbrich, Anja & Stanat, Petra (2015): Der Zusammenhang von sprachlichen und mathematischen Kompetenzen bei Kindern mit Deutsch als Zweitsprache. *Zeitschrift für Pädagogische Psychologie* 29 (1), 19–29. doi:https://doi.org/10.1024/1010-0652/a000142.

Serva, Maurizio & Petroni, Filippo (2008): Indo-European languages tree by Levenshtein distance. *EPL (Europhysics Letters)* 81 (6), 68005. doi:https://doi.org/10.1209/0295-5075/81/68005.

Sonnleitner, Philipp, Krämer, Charlotte, Gamo, Sylvie, Reichert, Monique, Muller, Claire, Keller, Ulrich & Ugen, Sonja (2018): Étude longitudinal des compétence des élèves. Évolution en compréhension écrite en allemand et en mathematiques entre la classe de 3e et la classe de 9e. In University of Luxemourg, Luxembourg Centre for Educational Testing & SCRIPT (eds.): *Rapport national sur l'Éducation au Luxembourg 2018*. Esch-sur-Alzette: University of Luxembourg, Luxembourg Centre for Educational Testing and SCRIPT. 39–58.

Sonnleitner, Philipp, Reichert, Monique & Ugen, Sonja (2014): Item development and test compilation. In Fischbach, Antoine, Ugen, Sonja, Martin, Romain (eds.): *ÉpStan Technical Report*. Luxembourg: University of Luxembourg, Luxembourg Centre for Educational Testing and SCRIPT. 5–11.

Vos, Teije de (1992): *Tempo-Test-Rekenen: test voor het vaststellen van het rekenvaardigheidsniveau der elementaire bewerkingen (automatisering) voor het basis- en voortgezet onderwijs: Handleiding*. Berkhout. "Swets & Zeitlinger"; Lisse.