

Brandenburg, Janin; Fischbach, Anne; Labuhn, Andju Sara; Rietz, Chantal Sabrina; Schmid, Johanna; Hasselhorn, Marcus

## Overidentification of learning disorders among language-minority students. Implications for the standardization of school achievement tests

*Journal for educational research online 8 (2016) 1, S. 42-65*



Quellenangabe/ Reference:

Brandenburg, Janin; Fischbach, Anne; Labuhn, Andju Sara; Rietz, Chantal Sabrina; Schmid, Johanna; Hasselhorn, Marcus: Overidentification of learning disorders among language-minority students. Implications for the standardization of school achievement tests - In: Journal for educational research online 8 (2016) 1, S. 42-65 - URN: urn:nbn:de:0111-pedocs-120293 - DOI: 10.25656/01:12029

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-120293>

<https://doi.org/10.25656/01:12029>

in Kooperation mit / in cooperation with:



**WAXMANN**  
[www.waxmann.com](http://www.waxmann.com)

<http://www.waxmann.com>

### Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### Kontakt / Contact:

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Janin Brandenburg, Anne Fischbach, Andju Sara Labuhn,  
Chantal Sabrina Rietz, Johanna Schmid & Marcus Hasselhorn

## Overidentification of learning disorders among language-minority students

Implications for the standardization of school achievement tests

### Abstract

*This German prevalence study examined disproportionate representation of language-minority students among children identified with learning disorder (LD) according to ICD-10 (WHO, 1992). Most German school achievement tests used in LD diagnostics do not provide separate norms for language-minority students, and thus do not take these children's second language status into account when evaluating their academic performance. Although this is likely to result in an LD overidentification of language-minority students, little is known about the magnitude of this effect. Therefore, we compared the estimation of LD prevalence between native German speaking students (n = 566) and language-minority students (n = 478) when pooled versus group-specific achievement norms were used for LD classification. Three important findings emerged from our study: Firstly, and as expected, significant disproportionality effects occurred under pooled norms. Specifically, the likelihood of being diagnosed with LD amounted to 14–18 % among native German speakers and nearly doubled to 25–30 % among language-minority students. Secondly, disproportionality varied as a function of LD subtype: Whereas no disproportionate representation was revealed for arithmetic LD (F81.2), overidentification of language-minority students was found for verbal LD subtypes (namely, reading disorder [F81.0], spelling disorder [F81.1],*

---

Dipl.-Psych. Janin Brandenburg (corresponding author) · Dr. Anne Fischbach · Dr. Andju Sara Labuhn · Dr. Chantal Sabrina Rietz · Prof. Dr. Marcus Hasselhorn, Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA) & German Institute for International Educational Research (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Germany

e-mail: brandenburg@dipf.de  
fischbach@dipf.de  
rietz@dipf.de  
hasselhorn@dipf.de

Dr. Johanna Schmid, Eberhard Karls University Tübingen, Institute of Psychology, Schleichstraße 4, 72076 Tübingen, Germany & German Institute for International Educational Research (DIPF), Schloßstraße 29, 60486 Frankfurt am Main, Germany  
e-mail: johanna.schmid@uni-tuebingen.de

and mixed disorder of scholastic skills [F81.3]). Thirdly, disproportionality effects were absent when group-specific norms were used for LD classification that controlled for second-language issues. Challenges that have to be met when testing language-minority students for LD are discussed.

### **Keywords**

Prevalence; Learning disorder; Language minority; Disproportionate representation

## **Überidentifikation von Lernstörungen bei Kindern mit Deutsch als Zweitsprache**

Implikationen für die Normierung von standardisierten Schulleistungstests

### **Zusammenfassung**

Die Prävalenzstudie untersucht bei Kindern, die Deutsch als Muttersprache (DaM) bzw. als Zweitsprache (DaZ) sprechen, die Häufigkeit von Lernstörungen nach ICD-10 (WHO, 1992). Die meisten deutschen Schulleistungstests, die zur Lernstörungsdiagnose herangezogen werden, stellen keine gesonderten Normen für Kinder mit DaZ bereit. Es ist anzunehmen, dass dies zu einer Überidentifikation von Lernstörungen bei Kindern mit DaZ führt, da die besondere Spracherwerbssituation dieser Kinder nicht berücksichtigt wird. Dennoch ist bislang wenig über das Ausmaß dieses Effektes bekannt. Die vorliegende Studie vergleicht daher die Lernstörungsprävalenz zwischen Drittklässlern mit DaM (n = 566) bzw. mit DaZ (n = 478) wenn gemeinsame versus getrennte Schulleistungsnormen zur Leistungsbeurteilung herangezogen werden. Die Studie erbrachte drei wesentliche Ergebnisse: (1) Wie erwartet kam es bei Verwendung gemeinsamer Schulleistungsnormen zu einer deutlichen Erhöhung der Lernstörungsprävalenz bei Kindern mit DaZ. Die Wahrscheinlichkeit einer Lernstörungsdiagnose belief sich für diese Teilstichprobe auf 25–30 % und war damit annähernd doppelt so groß wie bei Kindern mit DaM, für die sich eine Gesamtprävalenz von 14–18 % ergab. (2) Die Gruppenunterschiede variierten dabei in Abhängigkeit des Lernörungstypus: Während keine signifikant unterschiedlichen Prävalenzraten für die isolierte Rechenstörung (F81.2) nachweisbar waren, zeigten sich für die verbalen Lernörungstypen (d. h. Lese-Rechtschreibstörung [F81.0], isolierte Rechtschreibstörung [F81.1] und kombinierte Störung schulischer Fertigkeiten [F81.3]) signifikant erhöhte Prävalenzraten für Kinder mit DaZ. (3) Werden hingegen getrennte Schulleistungsnormen zur Lernstörungsdiagnose herangezogen um für die besondere Spracherwerbssituation von Kindern mit DaZ zu kontrollieren, nähern sich die Prävalenzraten beider Gruppen wie erwartet auf ein vergleichbares Niveau an. Es wird diskutiert, welche Herausforderungen sich bei der Lernstörungsdiagnostik von Kindern mit DaZ ergeben.

## Schlagworte

Prävalenz; Lernstörung; Deutsch als Zweitsprache; Testfairness

### 1. Introduction

Similar to many other countries, Germany's population is becoming increasingly diverse. In fact, the current German National Report on Education (Authoring Group Educational Reporting, 2014) pointed out that approximately 34 % of elementary school children in Germany have a migration background, and that this number is steadily increasing. Consequently, many children for whom the language of academic instruction is not their native language are entering the educational system. Stimulated by these developments, much research attention in recent years has been devoted to the academic success of language-minority students (also referred to as *non-native speakers* throughout this study). Overall, this branch of research consistently points towards the enormous challenges these children face in educational settings: For instance, language-minority students constantly lack behind their native German speaking classmates in academic achievement (e.g., mathematics: Heinze, Herwartz-Emden, & Reiss, 2007; reading: Schwippert, Wendt, & Tarelli, 2012; Stanat, Rauch, & Segeritz, 2010) and have double the risk of dropping out of school (Authoring Group Educational Reporting, 2014). Moreover, as international large-scale assessments like the *Programme for International Student Assessment* (PISA) have shown, particularly in Germany language-minority students are at a greater risk of failure in school (e.g., Organisation for Economic Co-operation and Development [OECD], 2003; Prenzel et al., 2007; Stanat et al., 2010). These alarming results have raised attention to the learning needs of the many children for whom the language spoken in school is not the native language.

From an intervention perspective, it is often crucial to identify the particular reasons why a child struggles in school and to examine whether or not the achievement problems are caused by an underlying learning disorder (LD). Yet, with respect to language-minority students, clarifying this question poses some diagnostic challenges to educational practitioners: Since the language of instruction is not the children's native language, poor academic achievement could primarily be caused by second language acquisition rather than by an underlying LD (cf. Orosco, Almanza de Schonewise, de Onis, Klingner, & Hoover, 2008). On the other hand, limited proficiency in the second language could also mask an LD so that it might be easily overlooked by diagnosticians (cf. Cline & Frederickson, 1999; Solari, Petscher, & Sidler Folsom, 2014). The difficulty of teasing apart these two possibilities may hamper a correct identification of affected children and, in turn, may delay the provision of appropriate support (cf. Samson & Lesaux, 2009). In this study we argue, that – in order to overcome this obstacle – special attention must be given to the normative samples of the school achievement tests used for

LD classification. Otherwise conclusions derived from test scores may be invalid and a reliable diagnosis of LD in language-minority students might not be possible.

### 1.1 Definition and diagnostics of learning disorders

In European countries, diagnostics of LD is usually based on the *International Classification of Diseases* (ICD-10, World Health Organization [WHO], 1992), according to which LDs are subsumed under the term *specific developmental disorders of scholastic skills*. Among the subtypes listed are specific reading disorder (F81.0), specific spelling disorder (F81.1), specific disorder of arithmetical skills (F81.2), and mixed disorder of scholastic skills (F81.3). The main feature of this category of disorders is a significant and unexpected impairment in the development of reading, spelling, and/or mathematical achievement: The learning problems are *significant* in that the child's performance is substantially below the level expected for the child's age and years of adequate schooling; and they are *unexpected* because they are in contradiction to the child's intellectual potential and are thus not a consequence of low IQ. This uncoupling between intelligence and academic achievement is at the heart of the medical definition of LDs (Ferrer, Shaywitz, Holahan, Marchione, & Shaywitz, 2010) and has led to the assumption that a (neuro-)biological dysfunction causes the learning problems experienced by children with LD. That is, LDs defined in this way are considered to be of biological origin and are thus distinguished from learning difficulties that are due to environmental or other factors such as inadequate opportunities to learn, emotional disturbances or negative family circumstance (cf. WHO, 1992).

LDs are operationalized in ICD-10 by the *double discrepancy criterion*, which requires that the child performs (a) below average compared to same-grade peers and also (b) below the level expected on the basis of his or her intelligence. According to the diagnostic guidelines of ICD-10 (WHO, 1992), the IQ-achievement discrepancy (i.e., criterion b) should be calculated using a regression formula which corrects for the imperfect correlation between IQ and achievement. By this means this formula also controls for regression to the mean.

### 1.2 Overidentification of learning disorders among language-minority students

Since LDs are clinical diagnoses caused by a biological dysfunction, there is neither a theoretical nor an empirical reason to expect that language-minority students should have a higher prevalence of LD than their native speaking classmates (see also Samson & Lesaux, 2009). Obviously, there is no neurobiological underpinning of whether a child speaks the language of instruction as the first or the additional language, and therefore prevalence rates should be equally distributed among both groups of children. However, there is some evidence to suggest that

the assessments and diagnostic procedures currently used in LD classification are negatively biased towards language-minority students, which in turn leads to a *disproportionate representation* of these children in special education programs (cf. Shifrer, Muller, & Calahan, 2011). Disproportionate representation exists when group membership influences the likelihood of being placed in a disability category (cf. Oswald, Coutinho, Best, & Singh, 1999). For instance, using a large U.S. American sample, Shifrer et al. (2011) demonstrated an alarming overrepresentation of language minorities among children who were diagnosed with reading disorder. Likewise, Artiles, Rueda, Salazar, and Higuera (2005) emphasized that a considerable number of language-minority students, namely more than 10 % of what would be expected given their proportion in the U.S. school population, were placed in LD secondary programs. Findings of Samson and Lesaux (2009) as well as Artiles, Rueda, Salazar, and Higuera (2002) showed further that LD overrepresentation of language-minority students increases with grade level. Overall, these studies illustrate the disproportionate representation of language minorities among children diagnosed with LD.

However, whereas research on LD in language-minority students has a two-decade long tradition in the U.S., this topic is relatively under-investigated in European countries such as Germany. In fact, reliable data is missing as to whether current diagnostic and assessment procedures in Germany also lead to LD overidentification among these children. Limited knowledge stems mostly from the fact that previous epidemiological studies have often excluded children who do not speak German as their native language. These studies (e.g., Strehlow & Haffner, 2002) thus shed light exclusively on the prevalence of LD among native speaking children. In contrast, the few epidemiological studies that included children with a first language other than German (e.g., Fischbach et al., 2013; Wyschkon, Kohn, Ballaschk, & Esser, 2009) have not reported prevalence rates as a function of the children's language status, and therefore, too, do not provide any information of potential disproportionality effects in current LD classification.

It is yet reasonable to assume that disproportionality is also an issue in Germany: Most achievement tests available for LD diagnostics do not provide separate norms for language-minority students and therefore do not take the children's second language status into account when evaluating their academic attainment. Instead, achievement norms are generally pooled across native and non-native speaking children. Although this is consistent with respect to the learning standards set by the German educational system, it is also problematic due to the following reason:

Unlike their monolingual classmates, language-minority students need to learn literacy and mathematics in a language that is not their native language. This is an enormous challenge, and – at least as long as the children are still in the process of acquiring the language of instruction – reduced content learning is likely to occur (cf. Butler & Stevens, 2001; Kieffer, Lesaux, Rivera, & Francis, 2009). These learning problems are a normal consequence of second language acquisition and should not be ascribed to a mental disorder such as an LD (cf. García, McKoon, & August,

2008). When evaluating children's academic performance as part of an LD diagnostics, it is therefore inevitable to control for second language issues.

It follows that achievement norms pooled across native and non-native speaking children might not be linguistically fair for LD testing among language-minority students: Such norms may introduce a systematic bias, which, in turn, may result in a disproportionate classification of language-minority students with LD. The reason for this is straightforward: Language minorities as a subpopulation generally score somewhat lower on standardized achievement tests than their monolingual peers (García et al., 2008; García & Pearson, 1994). Thus, when combining performance scores of both subpopulations for standardization purposes, the language-minority students' raw score distribution is forced towards the native speakers' distribution, and as a result *expected* mean performance is higher under pooled norms than under specific norms. Performance of language-minority students as a subpopulation is thus overestimated (i.e., the groups' *actual* mean performance is lower). As a consequence, these children might be more likely to be seen as meeting LD cut-off scores than their native speaking classmates. However, that they do so is not because they have an actual higher risk of LD. Instead, it is an artefact caused by the fact that two subpopulations are combined into one although their raw score distributions do not completely overlap.

Although it is likely that achievement norms pooled from native and non-native speaking children should result in LD *overidentification* among language-minority students and LD *underidentification* among native speaking children, to our knowledge this has not been examined systematically and therefore not much is known about the size and magnitude of this effect. If pooled achievement norms are indeed one source responsible for the disproportionality effect, then applying group-specific norms should reduce this classification bias and should result in a comparable prevalence estimation of LD for language-minority students and their native speaking classmates.

### 1.3 Summary of the present study

Most German school achievement tests do not provide separate norms for language-minority students and thus do not take these children's second-language status into account. With respect to the German learning standards this is adequate. However, since achievement tests are not only used to assess students' achieved level of competencies, but are also used in LD diagnostics, this is likely to result in a systematic LD overidentification among language-minority students. Although this disproportionality effect is likely to occur, it has not been examined thoroughly. Hence, using a large population-based sample, we examined the extent to which prevalence rates differed between native and non-native speaking children when pooled versus specific achievement norms were used in LD diagnostics.

A further aspect by which the present study adds to previous research is that disproportionality issues were not only examined in the classification of reading

disorder, but also in spelling disorder, arithmetic disorder, and in the mixed disorder of scholastic skills. By this means, we aimed to clarify whether disproportionality varies as a function of LD subtype: Since it has been suggested that second language acquisition puts more constraints on verbal subjects than on learning mathematics (Schwippert, Bos, & Lankes, 2003), it seems plausible to assume that disproportionality effects are lower for arithmetic LD – although they may nevertheless occur. In particular, we addressed the following research questions:

1. Are language-minority students in Germany disproportionately identified as having an LD when applying the diagnostic criteria and methods currently used in LD diagnostics?
2. If so, does disproportionality differ as a function of LD subtype?
3. Do disproportionality effects disappear when group-specific norms accounting for second language acquisition are used for LD classification?

## 2. Method

### 2.1 Participants

For this prevalence study, we used a population-based sample, which included 1,170 children at the beginning of their third grade (48.5 % male). The children came from 32 elementary schools in Frankfurt am Main, the fifth biggest city in Germany. We chose to investigate LD classification among third-graders, because this is the age group in which LDs are most frequently diagnosed (Hasselhorn & Schuchardt, 2006). Further, as we intended to compare prevalence rates for native and non-native German speaking children, we decided to conduct this study in an urban area, where we expected to find a comparable amount of children for each of the two groups.

The teachers were asked to indicate for each child whether his or her first language was German or not. We used this information to classify the children as either *native German speakers*, *non-native German speakers* or *bilingual speakers* (i.e., children who speak German and another language as their native language). In our sample, 48 % ( $n = 566$ ) of the children were native German speakers, 41 % ( $n = 478$ ) of the children were classified as language-minority students, because they had a first language other than German and the remaining 11 % ( $n = 126$ ) were classified as bilingual. This proportion is largely in line with the official population statistics of the region (City of Frankfurt am Main, 2012), which state that 42 % of the children aged 6 to 10 years have a migration background.

For the prevalence estimation, only children who were native or non-native speakers were considered. Bilingual children were excluded from further analyses, because this subsample was too small to provide a valid estimation of LD prevalence among bilinguals. The final sample's mean age was 8 years and 8 months ( $SD = 0.43$ ). The children's mean IQ was in the normal range ( $M = 104.0$ ;

$SD = 15.54$ ) as were their academic achievement scores ( $T$ -score in reading:  $M = 52.2$ ;  $SD = 10.07$ ;  $T$ -score in spelling:  $M = 49.2$ ;  $SD = 10.15$ ;  $T$ -score in mathematics:  $M = 50.4$ ;  $SD = 10.28$ ).

## 2.2 Measures

*Reading.* A standardized German decoding speed test for first to fourth graders (WLLP; Küspert & Schneider, 1998) was administered to assess children's reading achievement. In this test, decoding speed is measured with a picture-word-matching procedure: Each of the 140 items consists of four pictures and one written word. The children's task is to identify the picture that corresponds to the written word. The children are given five minutes to solve as many items as possible. The dependent variable is the number of correctly marked pictures. Parallel-forms reliability of this measure is reported as  $\alpha = .93$ .

*Spelling.* We assessed children's spelling achievement with a standardized German spelling test for first to third graders (DERET 1-2+; Stock & Schneider, 2008). This test requires children to spell 52 dictated words embedded in a short narrative. The dependent variable is the number of spelling errors. Internal consistency of this measure is reported as  $\alpha = .92$ .

*Mathematics.* To assess mathematical achievement, a standardized test for second to third graders (DEMAT 2+; Krajewski, Liehm, & Schneider, 2004) was administered, which requires children to solve mathematical problems related to arithmetic, magnitude, and geometry. The test consists of 36 items, which are given under time constraints. The dependent variable is the number of items solved correctly. Internal consistency of this measure is reported as  $\alpha = .94$  for third-graders.

*Nonverbal intelligence.* To obtain an estimate of general cognitive ability, children completed the German version of the Columbia Mental Maturity Scale 1-3 (CMM 1-3; Schuck, Eggert, & Raatz, 1975). The CMM 1-3 is a nonverbal measure of inductive reasoning and was used as an indicator of fluid intelligence. This non-timed test consists of 50 items in total. The items are composed of five figures each, four of which can be classified together whereas the remaining figure does not match the others. The children's task is to identify the odd-one-out item. The dependent variable is the number of correctly marked items. The technical manual reports a respectable internal consistency of  $\alpha = .87$ .

## 2.3 Procedure

The assessment took place in elementary schools. Student research assistants (i.e., psychology students who were trained in LD diagnostics) administered the tests in a classroom setting on two school days. In one session, children completed the spelling, reading, and IQ test, which lasted up to 90 minutes. Mathematical perfor-

mance was assessed during a second session lasting up to 45 minutes. The maximum interval between the sessions was one week. All instructions were provided verbally in German. Test administration and scoring was carried out according to the test manuals. Parental informed written consent was obtained for all children prior to testing.

## 2.4 Classification of learning disorders

Classification of LDs was based on the double discrepancy criterion according to ICD-10. Yet, whereas the ICD-10 *Diagnostic Criteria for Research* (WHO, 1993) suggest applying a cut-off score of at least 2 *SDs* below expected performance level, less stringent criteria are considered appropriate in educational practice: In general, cut-off scores of about 1.2 *SDs* correspond to the diagnostic guidelines recommended for (Strehlow & Haffner, 2002) and most frequently used (Hasselhorn, Mähler, & Grube, 2008; Klicpera, Schabmann, & Gasteiger-Klicpera, 2010) in German educational and clinical settings. The cut-off scores applied in the present study were therefore set at 1.2 *SDs*. In this way, our sample best represented the subpopulation of school children in Germany commonly referred to as having an LD.

LD classification criteria were as follows: Children's nonverbal IQ was at least 70, and their low achievement score in reading, spelling and/or mathematics was (a) at least 1.2 *SDs* below the normed reference group's mean (i.e., percentile  $\leq 12$ ; *T*-score  $\leq 38$ ), and (b) at least 1.2 *SDs* below their achievement level expected for intelligence. To determine criterion (b) we used the regression approach rather than the difference score method, as recommended for individual diagnostics (cf. Schulte-Körne, Deimel, & Remschmidt, 2001; WHO, 1992). Hence, a child's expected achievement level was estimated by the following formula (cf. Marx, Weber, & Schneider, 2001):

$$\text{expected achievement level} = r_{\text{achievement, IQ}} * \text{IQ} + (1 - r_{\text{achievement, IQ}}) * 100 \quad (1)$$

where  $r_{\text{achievement, IQ}}$  is the correlation between IQ and achievement. The respective correlations obtained from our own sample are presented in Table 1.

Children whose *actual* achievement level was at least 1.2 *SDs* below their *expected* level were considered discrepant with respect to IQ. In the regression formula, this discrepancy criterion is expressed in terms of standard deviation units of the residuum (cf. Schulte-Körne et al., 2001):

$$\text{standard deviation of the residuum} = 12 * \sqrt{(1 - r_{\text{achievement, IQ}})^2} \quad (2)$$

Table 1: Spearman's rank correlations between nonverbal intelligence and achievement scores

	$r_{IQ, Reading}$	$r_{IQ, Spelling}$	$r_{IQ, Mathematics}$
Original Norms ( $N = 1,044$ )	.239	.318	.450
Pooled Norms ( $N = 1,044$ )	.240	.290	.455
Group-specific Norms			
Native speakers ( $n = 566$ )	.208	.245	.453
Non-native speakers ( $n = 478$ )	.183	.254	.340

Note. Correlations are based on  $T$ -scores ( $M = 50$ ;  $SD = 10$ ). Spearman's rank correlations instead of Pearson's correlations were computed due to the non-normality of the data. All correlation coefficients are significant at  $p < .05$ .

In particular, a *specific spelling disorder* (F81.1) was classified when the child's spelling score fulfilled the double discrepancy criterion outlined above, whereas the reading and the math score were at least in the normal range. Likewise, a *specific disorder of arithmetical skills* (F81.2) was classified when the child's math score fulfilled the double discrepancy criterion, whereas performances in reading and spelling were at least in the normal range. Further, according to ICD-10, children were classified as having a *specific reading disorder* (F81.0) when either (a) only their reading score or (b) both their reading *and* their spelling score fulfilled the double discrepancy criterion, whereas their mathematical performance was at least in the normal range. Finally, children were diagnosed with a *mixed disorder of scholastic skills* (F81.3) when they met the criteria of an arithmetic LD in addition to a reading and/or spelling disorder.

## 2.5 Standardization of raw scores

To examine the extent to which pooled versus group-specific norms may influence LD prevalence among native and non-native speaking children, we had to standardize the raw scores on our own sample. *Pooled norms* were obtained by standardizing the raw scores on the *whole* sample of native and non-native children ( $N = 1,044$ ). *Group-specific norms* were obtained by standardizing the raw scores separately on each subsample. Since the raw scores were not normally distributed as indicated by inspection of histograms and the Kolmogorov-Smirnov test (reading:  $D(1,044) = 0.06$ ,  $p < .001$ ; spelling:  $D(1,044) = 0.07$ ,  $p < .001$ ; mathematics:  $D(1,044) = 0.11$ ,  $p < .001$ ; intelligence:  $D(1,044) = 0.06$ ,  $p < .001$ ), standardization was based on the area transformation technique developed by McCall (cf. Lienert & Raatz, 1998). Whereas the three school achievement tests (reading, spelling, and mathematics) were co-normed on the whole sample as well as on the two subsamples, the IQ test was normed on the whole sample only.

Our decision to restandardize the IQ test in addition to the achievement tests was due to the following reason: A one-sample  $t$ -test indicated that our

sample's mean IQ of 104 differed significantly from the expected value of 100,  $t(1043) = 8.29, p < .001$ . This difference might be attributed to the Flynn effect, which describes the phenomenon that intelligence test scores rise within populations over time (Flynn, 1987). Moreover, as Kanaya and Ceci (2012) recently demonstrated, the Flynn effect has a negative impact on LD prevalence estimation, because systematically more children than in the initial normative sample achieve higher test scores. This in turn leads to an overidentification of LD, because the children meet the IQ-achievement discrepancy criterion more easily. Thus, in order to ensure that our estimation of LD prevalence is not artificially inflated due to the Flynn effect, we decided to restandardize the IQ test on the whole sample (see also Fischbach et al., 2013). Table 2 provides the intelligence and achievement scores of the two subsamples before and after standardization.

**Table 2:** Descriptive statistics of the two subsamples as a function of standardization

	Native speakers		Non-native speakers	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Original Norms				
Intelligence <sup>a</sup>	107.20	15.31	100.18	14.95
Reading <sup>b</sup>	54.92	9.97	48.96	9.17
Spelling <sup>b</sup>	51.73	10.30	46.24	9.12
Mathematics <sup>b</sup>	52.91	9.72	47.31	10.11
Pooled Norms				
Intelligence <sup>a</sup>	103.01	14.63	96.88	14.53
Reading <sup>b</sup>	52.68	9.97	46.82	9.08
Spelling <sup>b</sup>	52.40	9.99	47.06	8.93
Mathematics <sup>b</sup>	52.51	9.76	47.01	9.42
Group-specific Norms				
Intelligence <sup>a</sup>	103.01	14.63	96.88	14.53
Reading <sup>b</sup>	50.04	9.96	50.00	10.01
Spelling <sup>b</sup>	50.05	9.92	49.99	9.99
Mathematics <sup>b</sup>	50.01	9.99	50.01	9.94

*Note.* <sup>a</sup> IQ-score ( $M = 100; SD = 15$ ); <sup>b</sup> T-score ( $M = 50; SD = 10$ ).

### 3. Results

Descriptive statistics of the LD subtypes are provided in Table 3 as a function of the children's language status. The respective prevalence rates are presented in Table 4. Results displayed in the upper section refer to the *original norms* (i.e., norms provided in the test manuals), results in the middle section refer to the *pooled norms* (i.e., norms from our own sample with data combined from native and non-native speaking children), and results in the lower section refer to the *group-specific norms* (i.e., norms from our own sample computed separately for native and non-native speaking children). We used chi-square tests of independ-

ence to determine associations between language status (native vs. non-native) and LD diagnosis (yes vs. no). Statistically significant effects were followed up with effect sizes in terms of odds ratios.

Table 3: Descriptive statistics of the LD subtypes: Means (*M*) and standard deviations (*SD*) for classification measures as a function of group and standardization

	Native speakers							
	Reading Disorder		Spelling Disorder		Arithmetic Disorder		Mixed Disorder	
	<i>M</i>	( <i>SD</i> )	<i>M</i>	( <i>SD</i> )	<i>M</i>	( <i>SD</i> )	<i>M</i>	( <i>SD</i> )
	Original Norms							
Intelligence <sup>a</sup>	105.6	(16.58)	102.6	(13.61)	97.0	(9.80)	95.9	(10.62)
Reading <sup>b</sup>	34.8	(2.41)	48.1	(7.13)	51.8	(7.98)	41.9	(7.93)
Spelling <sup>b</sup>	40.4	(8.23)	34.5	(3.08)	47.2	(5.99)	34.3	(5.65)
Mathematics <sup>b</sup>	50.9	(7.26)	51.0	(7.00)	33.0	(2.65)	31.9	(3.35)
	Pooled Norms							
Intelligence <sup>a</sup>	104.0	(17.02)	99.1	(12.06)	98.1	(6.79)	96.3	(9.41)
Reading <sup>b</sup>	33.2	(3.21)	48.3	(7.02)	52.09	(6.19)	37.9	(7.98)
Spelling <sup>b</sup>	42.3	(10.07)	33.5	(4.16)	46.8	(4.72)	36.3	(8.34)
Mathematics <sup>b</sup>	50.4	(6.44)	51.3	(6.33)	34.6	(3.24)	32.3	(4.31)
	Group-specific Norms							
Intelligence <sup>a</sup>	104.0	(16.85)	98.0	(12.39)	97.7	(8.68)	96.5	(8.58)
Reading <sup>b</sup>	33.4	(3.92)	47.8	(5.13)	49.6	(5.83)	36.6	(7.37)
Spelling <sup>b</sup>	41.3	(9.59)	34.2	(2.74)	46.7	(5.04)	35.3	(7.05)
Mathematics <sup>b</sup>	48.3	(6.11)	47.0	(6.27)	34.3	(2.93)	31.5	(5.48)
	Non-native speakers							
	Reading Disorder		Spelling Disorder		Arithmetic Disorder		Mixed Disorder	
	<i>M</i>	( <i>SD</i> )	<i>M</i>	( <i>SD</i> )	<i>M</i>	( <i>SD</i> )	<i>M</i>	( <i>SD</i> )
	Original Norms							
Intelligence <sup>a</sup>	99.1	(12.17)	102.3	(11.97)	98.5	(13.92)	96.1	(13.19)
Reading <sup>b</sup>	34.1	(3.59)	46.2	(5.72)	49.2	(7.19)	38.5	(5.99)
Spelling <sup>b</sup>	37.5	(6.68)	34.9	(2.79)	45.9	(5.48)	34.6	(5.19)
Mathematics <sup>b</sup>	46.1	(7.71)	47.5	(6.47)	32.9	(2.97)	31.1	(3.44)
	Pooled Norms							
Intelligence <sup>a</sup>	98.4	(10.82)	100.2	(10.03)	102.0	(13.35)	95.0	(10.92)
Reading <sup>b</sup>	33.3	(4.01)	46.3	(4.87)	47.6	(6.33)	35.5	(5.87)
Spelling <sup>b</sup>	40.1	(7.52)	35.0	(3.58)	46.9	(5.26)	35.5	(6.57)
Mathematics <sup>b</sup>	46.7	(6.33)	46.6	(5.22)	33.9	(3.93)	30.8	(5.21)
	Group-specific Norms							
Intelligence <sup>a</sup>	99.3	(12.25)	99.2	(10.99)	96.2	(13.49)	93.1	(8.36)
Reading <sup>b</sup>	34.1	(3.78)	47.2	(5.01)	48.9	(6.87)	37.0	(7.30)
Spelling <sup>b</sup>	41.0	(7.63)	34.7	(3.49)	48.1	(5.68)	36.9	(8.21)
Mathematics <sup>b</sup>	47.8	(7.51)	49.1	(6.67)	33.9	(3.39)	32.6	(3.33)

Note. <sup>a</sup> IQ-score (*M* = 100; *SD* = 15); <sup>b</sup> *T*-score (*M* = 50; *SD* = 10).

### 3.1 Prevalence of learning disorders when original norms were used for classification

Among native German speakers, one in six children (18 %) fulfilled the diagnostic criteria of any one of the four LDs of ICD-10 when original norms were used for classification. In contrast, the respective prevalence rate of the language-minority students was much higher: 30 %, that is one in three children, met the diagnostic criteria of any one of the four LD subtypes when original norms were used for classification. This difference in overall prevalence was statistically significant,  $\chi^2(1, N = 1,044) = 21.77, p < .001$ . The odds ratio revealed that the odds of being diagnosed with LD were two times higher for language-minority students than for native speaking children, odds ratio = 1.99 [95 % CI: 1.47–2.69].

Next, prevalence rates were examined separately for each of the four LD subtypes: In each domain, the prevalence was higher for language-minority students than for native speaking children. Yet, these differences were statistically significant only for spelling disorder,  $\chi^2(1, N = 1,044) = 6.58, p = .010$ , odds ratio = 1.86 [95 % CI: 1.12–3.08], and for the mixed disorder of scholastic skills,  $\chi^2(1, N = 1,044) = 13.62, p < .001$ , odds ratio = 3.12 [95 % CI: 1.60–6.16]. That is, when compared to their native speaking classmates, language-minority students were about twice as likely to fulfill the operational criteria of a spelling disorder, and were over three times as likely to fulfill the diagnostic criteria of a mixed disorder of scholastic skills. No statistically significant group differences were found for reading disorder,  $\chi^2(1, N = 1,044) = 1.43, p = .232$ ; and arithmetic disorder,  $\chi^2(1, N = 1,044) = 1.22, p = .270$ .

### 3.2 Prevalence of learning disorders when pooled norms were used for classification

Next, we examined how the composition of the test norms influences the estimation of LD prevalence. Specifically, using *pooled* norms we expected to replicate the disproportionality effect found under *original* norms, whereas under *group-specific* norms we rather expected to find no disproportionality effects at all. Although prevalence rates decreased in magnitude, this was exactly the pattern we found: Among native German speakers, one in seven children (14 %) fulfilled the diagnostic criteria of any one of the four LDs when pooled norms were used for classification. As previously found under original norms, this prevalence rate increased by approximately 11 % in language-minority students. Of them, one in five children met the diagnostic criteria of any one of the four LD subtypes. Again, this difference in overall prevalence was statistically significant,  $\chi^2(1, N = 1,044) = 18.75, p < .001$ , odds ratio = 1.99 [95 % CI: 1.44–2.76].

Table 4: Prevalence rates (in %) as a function of group and standardization

LD Subtype	Native speakers (n = 566)			Non-native speakers (n = 478)		
	n	m / f	%	n	m / f	%
Original Norms						
Reading Disorder <sup>a</sup>	23	16 / 7	4.1	27	17 / 10	5.6
Spelling Disorder	30	25 / 5	5.3	45	28 / 17	9.4
Arithmetic Disorder	33	8 / 24 <sup>b</sup>	5.8	36	9 / 27	7.5
Mixed Disorder	14	7 / 7	2.5	35	10 / 25	7.3
Total	100	56 / 43 <sup>b</sup>	17.7	143	64 / 79	29.8
Pooled Norms						
Reading Disorder	33	24 / 9	5.8	45	25 / 19 <sup>b</sup>	9.4
Spelling Disorder	16	15 / 1	2.8	25	17 / 8	5.2
Arithmetic Disorder	19	5 / 14	3.4	24	5 / 19	5.0
Mixed Disorder	12	5 / 7	2.1	24	7 / 17	5.0
Total	80	49 / 31	14.1	118	54 / 63 <sup>b</sup>	24.6
Group-Specific Norms						
Reading Disorder	48	33 / 15	8.5	40	20 / 19 <sup>b</sup>	8.4
Spelling Disorder	22	17 / 5	3.9	18	11 / 7	3.8
Arithmetic Disorder	27	4 / 23	4.8	17	4 / 13	3.6
Mixed Disorder	24	11 / 13	4.2	17	6 / 11	3.6
Total	121	65 / 56	21.4	92	41 / 50 <sup>b</sup>	19.4

*Note.* LD = Learning Disorder; m = male; f = female. <sup>a</sup> Whereas under original norms the lowest prevalence rates were found for reading disorder, this LD subtype showed the highest prevalence rates under pooled norms. This difference is due to the restandardization: Our final sample's mean reading score of  $T = 52$  (original norms) differed significantly from the expected value of 50. Thus, by restandardization our samples' raw scores were lowered to a mean value of 50. Hence, more children than under original norms fulfilled the low achievement criteria of  $PR \leq 12$ . In contrast, this effect did not occur for the spelling and mathematical scores. This is why the relative order of prevalence rates differed across the LD subtypes between original and pooled norms. <sup>b</sup> Gender of one participant is unknown.

A similar result pattern also emerged when classification rates were examined separately for each of the four LD subtypes: Prevalence for the language-minority students was again significantly higher with respect to spelling disorder,  $\chi^2(1, N = 1,044) = 3.97, p = .046$ , odds ratio = 1.90 [95 % CI: 0.96–3.77], and the mixed disorder of scholastic skills,  $\chi^2(1, N = 1,044) = 6.55, p = .010$ , odds ratio = 2.44 [95 % CI: 1.15–5.23]. Yet, in contrast to the previous analysis, a further statistically significant group effect emerged for reading disorder,  $\chi^2(1, N = 1,044) = 4.82, p = .028$ , odds ratio = 1.68 [95 % CI: 1.03–2.75]. Overall, language-minority students were 1.5 up to 2.5 times more likely to fulfill the diagnostic criteria of verbal LDs than their native speaking classmates. Again, no statistically significant group difference was found for arithmetic disorder,  $\chi^2(1, N = 1,044) = 1.82, p = .178$ . To summarize, using pooled norms from our own sample, we were largely able to replicate the findings obtained when LD classification was based on the original test norms. These results support the notion that disproportionality issues in LD diagnostics occur when language-minority student's academic achievement is evaluated against normative samples that include a lar-

ge proportion of children who speak the language of instruction as their first language.

### **3.3 Prevalence of learning disorders when group-specific norms were used for classification**

Lastly, we examined differences in prevalence rates when group-specific norms were used for LD classification. By this means, we controlled for learning problems related to second language acquisition. We hypothesized that group-specific norms reduce classification bias, which should result in a comparable estimation of LD prevalence among native and non-native speaking children. Thus, in this particular case, the null hypothesis was tested. We therefore set the alpha level at  $p < .10$  instead of using the conventional level of  $.05$  (Bortz & Schuster, 2010). Modifying the alpha level in this way increases statistical power and reduces the probability that the null hypothesis is accepted although it is false (Type II error).

When group-specific norms in reading, spelling, and mathematics were used for classification, 21 % of the native speakers and 19 % of the non-native speakers fulfilled the diagnostic criteria of any one of the four LD subtypes described in ICD-10. Even under the increased alpha level of  $.10$ , this difference in overall prevalence rate was statistically non-significant,  $\chi^2(1, N = 1,044) < 1, p = .395$ . That is, a comparable number of native and non-native speaking children, namely one child in five, was classified with an LD. The same pattern of results was revealed for each of the four LD subtypes: Estimated prevalence rates were almost identical between the two groups and did not differ statistically (reading disorder:  $\chi^2(1, N = 1,044) < 1, p = .948$ ; spelling disorder:  $\chi^2(1, N = 1,044) < 1, p = .919$ ; arithmetic disorder:  $\chi^2(1, N = 1,044) < 1, p = .331$ ; mixed disorder:  $\chi^2(1, N = 1,044) < 1, p = .571$ ).

## **4. Discussion**

Recent research – mostly conducted in the U.S. – has drawn attention to the issue that the assessments and diagnostic procedures currently used in LD classification might be negatively biased towards language-minority students (e.g., Shiffrer et al., 2011). As a consequence, these students are often disproportionately represented among children identified with an LD (e.g., Artiles et al., 2002; 2005). Up to now, reliable data is missing as to whether such disproportionality issues are also present in Germany. This is because previous epidemiological studies have either excluded language-minority students from prevalence estimation or have not reported prevalence rates as a function of the children's language status.

Therefore, the first objective of this prevalence study was to examine whether the risk of LD classification differs between native and non-native speaking chil-

dren in Germany. To this end, 566 native and 478 non-native German speaking children completed a set of German school achievement tests that are commonly used in LD diagnostics. Based on the standard scores provided in the respective test manuals, LD prevalence was determined separately for each of the two groups and statistically compared to each other.

Next, we examined the extent to which the composition of the normative sample may introduce disproportionality effects in LD testing: German school achievement tests usually provide norms that are pooled across native as well as non-native speaking children; additional norms specifically for language-minority students are usually not provided. Clearly, those pooled norms are an essential and indispensable tool when teachers want to evaluate whether their students reach the German learning standards, which define what a child at a certain grade level should be capable of. Yet, those pooled norms might be problematic when testing language-minority students for LD: They may introduce a classification bias, because they do not take the children's second language status in account. Although it is likely that achievement norms pooled from native and non-native speaking children will result in an LD overidentification among language-minority students, not much is known about the size and magnitude of this effect. Therefore, the second objective of this study was to examine differences in prevalence rates when pooled norms versus group-specific norms were used for LD classification. To this end, we standardized the school achievement tests on our own sample and by this means derived pooled as well as group-specific norms. The underlying expectation was as follows: If pooled achievement norms are one source responsible for the disproportionality effect, then applying group-specific norms should reduce classification bias and should result in a comparable prevalence estimation of LD in native and non-native speaking children.

#### **4.1 Are language-minority students in Germany disproportionately identified as having a learning disorder when applying the diagnostic criteria and methods currently used in diagnostics?**

As expected, marked differences in prevalence rates were found between native and non-native speaking children. In particular, the likelihood of being diagnosed with LD amounted to 14–18 % among native speakers and nearly doubled to 25–30 % among language-minority students. Thus, when compared to their native speaking classmates, twice as many language-minority students were classified as having an LD when undifferentiated norms (i.e., original or pooled norms) were used for classification. Clearly, those high rates of overidentification are alarming. Since there is neither a theoretical nor an empirical reason to assume that LDs should occur more frequently in children speaking the language of instruction as a second language (cf. Samson & Lesaux, 2009), it is highly essential to draw attention to this problem and rethink current diagnostic methods used in LD testing.

## 4.2 Does disproportionality differ as a function of learning disorder subtype?

To the best of our knowledge, previous studies on disproportionate representation of LD did not report overidentification rates as a function of different LD subtypes. Therefore, little is known about whether LD overidentification varies depending on the learning domain. As we assessed performance in all *the three R's* (reading, writing, and arithmetic) it was possible to address this issue and thus examine disproportionality effects separately for each of the four LD subtypes.

Using original as well as pooled norms, the largest disproportionality effects were observed for the mixed disorder of scholastic skills: Language-minority students were 2.5 up to 3.1 times more likely to fulfil the diagnostic criteria of this LD subtype than their native German speaking classmates. Thus, when undifferentiated norms are used, language-minority students are especially at risk of being labelled with a mixed disorder of scholastic skills, which is commonly regarded as the most severe LD subtype (Gathercole & Alloway, 2008).

Comparably large effects of overidentification were found for spelling disorder: Under both original and pooled norms, language-minority students showed prevalence rates that were approximately twice as high as the ones obtained for their native speaking classmates. In contrast, findings related to reading disorder were not as clear: Whereas no significant disproportionality effect was found under original norms, a 1.5 times higher risk ratio to the disadvantage of language-minority students occurred under pooled norms. In order to better understand this effect, we compared the classification results at the individual level. This analysis revealed that some of the language-minority students who were previously classified with a mixed disorder under original norms were reclassified with reading disorder under pooled norms, because they now failed to meet the critical discrepancy criterion with respect to mathematics. In contrast, this reclassification from a mixed disorder under original norms to a reading disorder under pooled norms did not emerge for any of the native speaking children. This may explain why the group difference with respect to reading disorder was more pronounced under pooled norms than under original norms.

The finding that disproportionality was a larger issue in spelling disorder than in reading disorder might be attributed to the fact that in the German language, learning to spell is generally more demanding than learning to read (cf. Wimmer & Mayringer, 2002). This phenomenon is due to the fact that orthographic regularity in the German language is true only for grapheme-to-phoneme correspondence (relevant in reading), but not for phoneme-to-grapheme correspondence (relevant in spelling).

In contrast, no disproportionality effects were revealed with respect to the arithmetic LD: Estimated prevalence rates were similar across native and non-native speaking children. This pattern of results was found for original as well as pooled norms. Thus, whereas national and international comparative studies (e.g., Dummert, Endlich, Schneider, & Schwenck, 2014; Stanat et al., 2010) demonstra-

ted that language-minority students in Germany struggle to a greater extent in mathematics than their native speaking classmates, our results show that they are not yet at increased risk of being diagnosed with an arithmetic LD. In a broader sense, this finding may suggest that the underlying factors that are associated with a specific LD in mathematics are relatively independent from those associated with second language acquisition. Nevertheless, as our study focussed on the lower achievement level only, this result does not necessarily mean that proficiency in the language of instruction is altogether unimportant when it comes to learning mathematics *in general*. For example, investigating the *full range* of mathematical performance, Heinze et al. (2007) showed that second language competencies were significantly related to mathematical achievement in children with migration background. In fact, there is some evidence to suggest that the cognitive factors that are associated with an arithmetic LD are not necessarily identical to those associated with learning mathematics in general (e.g., Grube & Seitz-Stein, 2012). Clearly, there is a need for further studies investigating the specific relationship between second language learning and mathematics, in both the full and the lower range of skill performance.

To summarize, whereas overidentification of language-minority students was evident in all those LD subtypes that included the verbal domain, no differences in prevalence rates were found for the specific disorder of arithmetical skills. In addition, prevalence rates were mostly similar irrespective of whether original or pooled norms were used for classification.

### **4.3 Do disproportionality effects disappear when group-specific norms accounting for second language acquisition are used for classification of learning disorders?**

When we used group-specific rather than pooled norms to evaluate language-minority students' academic achievement, their overall prevalence rate decreased by a fifth, namely from 25 % to 19 %. In contrast, among native German speaking children the respective prevalence rate increased from 14 % to 21 %. That is, under group-specific norms prevalence rates of the two subsamples converged up to a statistically comparable percentage, and did not differ anymore. The same pattern of results was found for each of the four LD subtypes. This finding is in line with the notion that combining two subsamples with different performance levels in order to form a common norm, may lead to a systematic *overidentification* of one of the subsamples and to an *underidentification* of the other (cf. Share & Silvia, 2003).

With respect to second language issues, these results clearly indicate that we have to ensure that the diagnostic instruments used in LD decision making are culturally and linguistically fair (cf. Cline, 2000; Cline & Frederickson, 1999). In fact, the just-released ICD-11 beta draft includes some proposals for changes to the current version of ICD-10. Of relevance to the present issue, it is proposed to include the notion that LD is not due to a lack of proficiency in the language of instruction (WHO, 2015). That is, it needs to be acknowledged that, in contrast to their

native speaking classmates, language-minority students face the additional challenge of learning in a language that is not their native language, which may reduce the efficiency of content learning (cf. Abedi & Gándara, 2006). A valid diagnosis of LD among language-minority students can thus only be achieved when we take the children's second language status into account (cf. García et al., 2008). As our study suggests, one possibility of accounting for these language issues is to evaluate the child's scholastic skills against the performance level expected for other children who also acquire the language of instruction as a second language.

#### **4.4 Limitations of the study**

There are at least two limitations to this study: First, our classification of language status was based on the information provided by each child's elementary school teacher who answered the question whether or not the child's native language was German. Clearly, this information is rather superficial since it does not include important facts such as how old the child was when the family came to Germany (i.e., his/her age of first exposure to the German language; his/her length of exposure; Schulz, 2013), or how proficient the child or his/her parents were in the German language.

Second, we examined disproportionate representation in a population-based sample only. It is reasonable to assume that the underlying mechanisms are even more complex in a clinical sample, because additional social and environmental factors may come into play. For instance, in Germany little is known about potential differences in LD referral: Hence, in future studies it would be important to clarify whether teachers refer language-minority students more or less often for LD testing than their native speaking students. Addressing this issue in a Canadian sample, Limbos and Geva (2001), for example, found that the elementary school teachers in this study showed low sensitivity in identifying language-minority students at risk for LD, which resulted in less referral for remedial services. If similar differences in LD referral also exist in Germany, it may be another aspect that further complicates the disproportionality issue.

#### **4.5 Practical implications of the study and directions for future research**

Our study has some crucial implications. The main conclusion derived from this data is: The composition of the normative sample matters! In order to validly diagnose an LD in language-minority students it is inevitable to take the children's second language status into account. As our results suggest, one way to do this is to use group-specific norms where available: This is especially true when evaluating the children's performance in reading and spelling, and is less urgent when examining their mathematical skills.

Likewise, our findings also indicate that *native* German speaking children are at increased risk of being *underidentified* with LD when pooled norms are used for classification. Underidentification is as much a problem as overidentification, because underidentified children are at risk of not receiving the special learning support they may need. A solution to this problem might be to use group-specific norms for LD testing in native speakers as well.

It is nevertheless worth mentioning that group-specific norms as well as the double discrepancy criterion should not be applied always and exclusively. Rather, it is the diagnostic question at hand determining which of the reference groups (i.e., pooled norms versus group-specific norms) and which definition of learning problems is the best one to choose. For example, if the objective is to quantify the extent to which a particular child, whether native German speaking or not, reaches the federal learning standard set by the German educational system, it seems reasonable to put the child's performance in relation to the total sample of same-grade learners. Pooled norms in combination with a single discrepancy criterion (i.e., grade-level discrepancy only) should thus be used for educational evaluations, as they provide the best information as to whether the child struggles in school and needs additional help with the learning content. However, if the diagnostic objective is to evaluate further whether those learning problems are caused by an underlying LD (i.e., neurobiological dysfunction according to ICD-10), then the double discrepancy criterion in combination with group-specific norms that account for second language status should be used for classification.

With respect to future research, a major challenge is to develop reading and spelling tests that include an additional set of norms for language-minority students. In fact, to date there is only a very limited number of German achievement tests available that provide those additional norms. Future studies should also determine whether norms that simply dichotomize between native versus non-native speaking children are sufficient or whether more fine-grind norms are needed in order to ensure a valid and fair LD diagnosis: Obviously, language-minority students are a very heterogeneous group. For example, these children differ from each other with respect to their length of exposure to the German language, which impacts on their language proficiency and as a consequence also on their content learning in school. Relatedly, with respect to the diagnosis of specific language impairments it has thus been claimed, that the language norms for language-minority students should not only be differentiated by age but also according to the length of exposure to the second language (Schulz, 2013). It remains to be seen, whether or not such fine-grind norms are also needed in LD diagnosis.

In sum, many challenges have to be met when determining LD among language-minority students. Specifically, diagnosticians and teachers have to disentangle whether given literacy difficulties in these students are primarily the result of an underlying LD or are due to limited contact with the language of instruction (Artiles et al., 2005; Klingner, Artiles, & Barletta, 2006; Shifrer et al., 2011; von Suchodoletz, 2007). This diagnostic challenge is further complicated by the fact that in Germany, no official recommendations are yet available as to how to

diagnose LD among language-minority students. As a first solution to this problem we emphasize that it is important to use group-specific norms when evaluating language-minority students' academic performance in the diagnostic process. At a time when the number of students from linguistically diverse backgrounds is steadily increasing among the school-aged population, addressing this issue of test fairness is of utmost importance.

## Acknowledgments

This study was prepared in the context of the Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), and was funded by the Hessian Initiative for the Development of Scientific and Economic Excellence (LOEWE).

## References

- Abedi, J., & Gándara, P. (2006). Performance of English language learners as a subgroup in large-scale assessments: Interaction of research and policy. *Educational Measurement: Issues and Practice*, 25(4), 36–46. doi:10.1111/j.1745-3992.2006.00077x
- Artiles, A. J., Rueda, R., Salazar, J. J., & Higuera, I. (2002). English-language learner representation in special education in California urban school districts. In D. J. Losen & G. Orfield (Eds.), *Racial inequality in special education* (pp. 117–136). Cambridge, MA: Harvard Education Press.
- Artiles, A. J., Rueda, R., Salazar, J. J., & Higuera, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children*, 71(3), 283–300. doi:10.1177/001440290507100305
- Authoring Group Educational Reporting. (2014). *Education in Germany 2014 – An indicator-based report including an analysis of people with disabilities in the education system*. doi:10.3278/6001820dw
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler [Statistics for human and social scientists]* (7th ed.). Berlin, Germany: Springer.
- Butler, F. A., & Stevens, R. (2001). Standardized assessment of the content knowledge of English language learners K-12: Current trends and old dilemmas. *Language Testing*, 18(4), 409–427. doi:10.1177/026553220101800406
- City of Frankfurt am Main, Germany – Stadtschulamt (2012). *Bildung in Frankfurt am Main, Bildungsbericht 2012 [Education in Frankfurt am Main, Report on Education 2012]*. Retrieved from: [https://www.frankfurt.de/sixcms/detail.php?id=2957&ffmpar\[\\_id\\_inhalt\]=9605163](https://www.frankfurt.de/sixcms/detail.php?id=2957&ffmpar[_id_inhalt]=9605163)
- Cline, T. (2000). Multilingualism and dyslexia: Challenges for research and practice. *Dyslexia*, 6(1), 3–12. doi:10.1002/(SICI)1099-0909(200001/03)6:1<3::AID-DYS156>3.0.CO;2-E
- Cline, T., & Frederickson, N. (1999). Identification and assessment of dyslexia in bi/multilingual children. *International Journal of Bilingual Education and Bilingualism*, 2(2), 81–93. doi:10.1080/13670059908667681
- Dummert, F., Endlich, D., Schneider, W., & Schwenck, C. (2014). Entwicklung schriftsprachlicher und mathematischer Leistungen bei Kindern mit und ohne Migrationshintergrund [Development of reading, spelling, and mathematics achieve-

- ment in children with or without a migration background]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 46(3), 115–132. doi:10.1026/0049-8637/a000110
- Ferrer, E., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2010). Uncoupling of reading and IQ over time: Empirical evidence for a definition of dyslexia. *Psychological Science*, 21(1), 93–101. doi:10.1177/0956797609354084
- Fischbach, A., Schuchardt, K., Brandenburg, J., Kleszczewski, J., Balke-Melcher, C., Schmidt, C., Büttner, G., Grube, D., Mähler, C., & Hasselhorn, M. (2013). Prävalenz von Lernschwächen und Lernstörungen: Zur Bedeutung der Diagnosekriterien [Prevalence of poor learners and children with learning disorders: Investigating the role of diagnostic criteria]. *Lernen und Lernstörungen*, 2(2), 65–76. doi:10.1024/2235-0977/a000035
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101(2), 171–191. doi:10.1037/0033-2909.101.2.171
- García, G. E., McKoon, G., & August, D. (2008). Language and literacy assessment. In D. August & T. Shanahan (Eds.), *Developing reading and writing in second-language learners: Lessons from the report of the National Literacy Panel on language-minority children and youth* (pp. 251–274). New York, NY: Routledge.
- García, G. E., & Pearson, P. D. (1994). Assessment and diversity. In L. Darling-Hammond (Ed.), *Review of research in education* (Vol. 20, pp. 337–392). Washington, DC: American Educational Research Association.
- Gathercole, S. E., & Alloway, T. P. (2008). *Working memory and learning: A practical guide for teachers*. London, United Kingdom: Sage.
- Grube, D., & Seitz-Stein, K. (2012). Arbeitsgedächtnis und Rechnen [Working Memory and Arithmetic]. In M. Hasselhorn & C. Zoelch (Eds.), *Funktionsdiagnostik des Arbeitsgedächtnisses* (pp. 145–157). Göttingen, Germany: Hogrefe.
- Hasselhorn, M., Mähler, C., & Grube, D. (2008). Lernstörungen in Teilleistungsbereichen [Domain-specific learning disabilities]. In R. Oerter & L. Montada (Eds.), *Entwicklungspsychologie* (pp. 769–778). Weinheim, Germany: Psychologie Verlagsunion.
- Hasselhorn, M., & Schuchardt, K. (2006). Lernstörungen: Eine kritische Skizze zur Epidemiologie [Learning disabilities: A critical sketch on epidemiology]. *Kindheit und Entwicklung*, 15(4), 208–215. doi:10.1026/0942-5403.15.4.208
- Heinze, A., Herwartz-Emden, L., & Reiss, K. (2007). Mathematikkenntnisse und sprachliche Kompetenz bei Kindern mit Migrationshintergrund zu Beginn der Grundschulzeit [Mathematical ability and language competence among children with migration background at the beginning of elementary school]. *Zeitschrift für Pädagogik*, 53, 562–581.
- Kanaya, T., & Ceci, S. (2012). The impact of the Flynn effect on LD diagnoses in special education. *Journal of Learning Disabilities*, 45(4), 319–326. doi:10.1177/0022219410392044
- Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English Language Learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research*, 79(3), 1168–1201. doi:10.3102/0034654309332490
- Klicpera, C., Schabmann, A., & Gasteiger-Klicpera, B. (2010). Definition, Häufigkeit und Prognose von Lese- und Rechtschreibschwierigkeiten [Definition, frequency, and prognosis of reading and spelling difficulties]. In C. Klicpera, A. Schabmann, & B. Gasteiger-Klicpera (Eds.), *Legasthenie – LRS. Modelle, Diagnose, Therapie und Förderung* (pp. 124–136). München, Germany: Ernst Reinhardt.
- Klingner, J. K., Artiles, A. J., & Barletta, L. M. (2006). English language learners who struggle with reading: Language acquisition or LD? *Journal of Learning Disabilities*, 39(2), 108–128. doi:10.1177/00222194060390020101

- Krajewski, K., Liehm, S., & Schneider, W. (2004). *Deutscher Mathematiktest für zweite Klassen (DEMAT 2+)* [German test of mathematical skills for second grade]. Göttingen, Germany: Hogrefe.
- Küspert, P., & Schneider, W. (1998). *Würzburger Leise Leseprobe* [Würzburger's silent reading test]. Göttingen, Germany: Hogrefe.
- Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* [Test construction and test analysis]. Weinheim, Germany: Beltz.
- Limbos, M., & Geva, E. (2001). Accuracy of teacher assessments of second-language students at risk for reading disability. *Journal of Learning Disabilities*, 34(2), 137–151. doi:10.1177/002221940103400204
- Marx, P., Weber, J.-M., & Schneider, W. (2001) Legasthenie versus allgemeine Lese-Rechtschreibschwäche: Ein Vergleich der Leistungen in der phonologischen und visuellen Informationsverarbeitung [Dyslexia versus garden-variety poor reading: A comparison of phonological and visual processing skills]. *Zeitschrift für Pädagogische Psychologie*, 15(2), 85–98. doi:10.1024//1010-0652.15.2.85
- OECD – Organisation for Economic Co-operation and Development (2003). *Where immigrant students succeed – A comparative review of performance and engagement*. Paris, France: OECD.
- Orosco, M. J., Almanza de Schonewise, E., de Onis, C., Klingner, J. K., & Hoover, J. J. (2008). Distinguishing between language acquisition and learning disabilities among English language learners. In J. K. Klingner, J. J. Hoover, & L. M. Baca (Eds.), *Why do English language learners struggle with reading?* (pp. 5–16). Thousand Oaks, CA: Corwin Press.
- Oswald, D. P., Coutinho, M. J., Best, A. M., & Singh, N. N. (1999). Ethnic representation in special education: The influence of school-related economic and demographic variables. *Journal of Special Education*, 32(4), 194–206. doi:10.1177/002246699903200401
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., & Pekrun, R. (2007). *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* [PISA 2006. Results of the third international comparative study]. Münster, Germany: Waxmann.
- Samson, J. F., & Lesaux, N. K. (2009). Language-minority learners in special education: Rates and predictors of identification for services. *Journal of Learning Disabilities*, 42(2), 148–162. doi:10.1177/0022219408326221
- Schuck, K.-D., Eggert, D., & Raatz, U. (1975). *Columbia Mental Maturity Scale (CMM 1–3)*. Weinheim, Germany: Beltz.
- Schulte-Körne, G., Deimel, W., & Remschmidt, H. (2001). Zur Diagnostik der Lese-Rechtschreibstörung [On the diagnostics of reading and spelling disorder]. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 29(2), 133–116.
- Schulz, P. (2013). Sprachdiagnostik bei mehrsprachigen Kindern [Language assessment in multilingual children]. *Sprache, Stimme, Gehör*, 37(4), 191–195. doi:10.1055/s-0033-1358700
- Schwippert, K., Bos, W., & Lankes, E.-M. (2003). Heterogenität und Chancengleichheit am Ende der vierten Klassenstufe im internationalen Vergleich [Heterogeneity and equal opportunities at the end of fourth grade – an international perspective]. In W. Bos, E.-M. Lankes, K. Prenzel, K. Schwippert, G. Walther, & R. Valtin (Eds.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* [First results of IGLU. An international comparison of student performances at the end of fourth grade] (pp. 265–302). Münster, Germany: Waxmann.
- Schwippert, K., Wendt, H., & Tarelli, I. (2012). Lesekompetenzen von Schülerinnen und Schülern mit Migrationshintergrund [Reading competence of children with migration background]. In W. Bos, I. Tarelli, A. Bremerich-Vos, & K. Schwippert (Eds.),

- IGLU 2011: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich [IGLU 2011: Reading competencies of elementary school children in Germany in comparison to other countries] (pp. 191–207). Münster, Germany: Waxmann.
- Share, D. L., & Silvia, P. A. (2003). Gender bias in LD-discrepancy and post-discrepancy definitions of reading disability. *Journal of Learning Disabilities, 36*(1), 4–16. doi:10.1177/00222194030360010201
- Shifrer, D., Muller, C., & Calahan, R. (2011). Disproportionality and learning disabilities: Parsing apart race, socioeconomic status, and language. *Journal of Learning Disabilities, 44*(3), 246–257. doi:10.1177/0022219410374236
- Solari, E. J., Petscher, Y., & Sidler Folsom, J. (2014). Differentiating literacy growth of ELL students with LD from other high-risk subgroups and general education peers: Evidence from Grade 3–10. *Journal of Learning Disabilities, 47*(4), 329–348. doi:10.1177/0022219412463435
- Stanat, P., Rauch, D., & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund [Students with migration background]. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider, & P. Stanat (Eds.), *PISA 2009. Bilanz nach einem Jahrzehnt [PISA 2009. Conclusions after a decade]* (pp. 200–229). Münster, Germany: Waxmann.
- Stock, C., & Schneider, W. (2008). *Deutscher Rechtschreibtest für das erste und zweite Schuljahr (DERET 1-2+)* [German spelling test for grades 1 and 2]. Göttingen, Germany: Hogrefe.
- Strehlow, U., & Haffner, J. (2002). Definitionsmöglichkeiten und sich daraus ergebende Häufigkeit der umschriebenen Lese- bzw. Rechtschreibstörung – theoretische Überlegungen und empirische Befunde an einer repräsentativen Stichprobe junger Erwachsener [Alternative definitions and the resulting prevalence rates of spelling disorder – Theoretical considerations and empirical findings in an epidemiological sample of adolescents and young adults]. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 30*(2), 113–126. doi:10.1024//1422-4917.30.2.113
- von Suchodoletz, W. (2007). Lese-Rechtschreibstörung (LRS) im Sprachenvergleich und im Fremdsprachenunterricht [Dyslexia in different languages and the effect of dyslexia on foreign language learning]. *Sprache Stimme Gehör, 31*(3), 126–131. doi:10.1055/s-2007-985401
- Wimmer, H., & Mayringer, H. (2002). Dysfluent reading in the absence of spelling difficulties: A specific disability in regular orthographies. *Journal of Educational Psychology, 94*(2), 272–277. doi:10.1037//0022-0663.94.2.272
- WHO – World Health Organization (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines* (10th Rev. ed.). Geneva, Switzerland: WHO.
- WHO – World Health Organization (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research* (10th Rev. ed.). Geneva, Switzerland: WHO.
- WHO – World Health Organization (2015, September). *ICD-11 beta draft*. Retrieved from <http://apps.who.int/classifications/icd11>
- Wyschkon, A., Kohn, J., Ballaschk, K., & Esser, G. (2009). Sind Rechenstörungen genauso häufig wie Lese-Rechtschreibstörungen? [Is a specific disorder of arithmetic skills as common as reading/spelling disorder?] *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 37*(6), 499–512. doi:10.1024/1422-4917.37.6.499