Assessing Cognitive Ability in Multicultural and Multilingual Settings: A Language-Fair Test Battery for Children (Integral version)

Dzenita Kijamet, Rachel Wollschläger & Sonja Ugen | DOI: https://doi.org/10.48746/bb2024lu-en-17

Suggested citation for this document:
Kijamet, D., Wollschläger, R., & Ugen, S. (2024). *Assessing Cognitive Ability in Multicultural and Multilingual Settings: A Language-Fair Test Battery for Children (Integral version).* Luxembourg Centre for Educational Testing (LUCET) & Service de la Recherche et de l'Innovation pédagogiques (SCRIPT). https://doi.org/10.48746/bb2024lu-en-17

This document appears as supplementary material in connection with the Luxembourg National Education Report 2024 – specifically as an integral version of the following article:

Kognitive Fähigkeiten im multikulturellen und mehrsprachigen Umfeld: Ein sprachlich faires Testverfahren für Kinder
Dzenita Kijamet, Rachel Wollschläger & Sonja Ugen

The "Luxembourg National Education Report 2024" is published in German and French and can be accessed at the following link:

www.bildungsbericht.lu

Assessing Cognitive Ability in Multicultural and Multilingual Settings: A Language-Fair Test Battery for Children (Integral version)

In the field of educational counselling and intervention, intelligence tests are widely used to help assess children for a specific learning disorder or identify giftedness, under- or overachievement. Numerous studies led to the widespread finding that intelligence is one of the most reliable indicators of academic and professional success (e.g., Kuncel et al., 2004; Neisser et al., 1996; Roth et al., 2015). However, it is essential to note that most of these tests will only be reliable and valid if conducted in a language the child is proficient in. This represents a considerable challenge in a highly multicultural and multilingual setting like Luxembourg. Particularly as practitioners must rely on tests from other countries that have not been adapted to the Luxembourgish setting in a standardized manner due to a lack of alternatives. Children who speak a different native language than the test language are often at a disadvantage (Ugen et al., 2021), and thus, relying on language-specific diagnostic tools can produce biased results in children who are not proficient in that language resulting in invalid test results.

Since 1905, researchers have been aware that language and culture can influence performance in cognitive tests (Schaap, 2011), leading to the development of nonverbal intelligence tests (e.g., using culture- and language-free content such as abstract figures). However, the instructions are still often language-based and nonverbal communication (e.g., using pantomime or gestures) can be unfair, as it is not necessarily clearer for participants and cultural meanings behind gestures vary.

In this chapter, we will present a test battery assessing cognitive ability designed to be a language-fair assessment tool for grade 3 children in multilingual settings. We will present the process of test development, illustrate one of the subtests in more detail and give a first glimpse into the results regarding language fairness.

Challenges of multilingualism and why a new test battery is necessary

In a multilingual educational setting such as in Luxembourg, where the main teaching language changes over the course of the educational curriculum and the school population is linguistically highly diverse, the availability of a standardized language-fair cognitive test battery is thus essential to adequately measure the children's cognitive potential in non-scholastic tasks (such as spatial relations, short-term memory etc.). In elementary school, for instance, the main teaching language in most public schools is German, which is linguistically close to Luxembourgish. However, only a third of the pupils speak German or Luxembourgish at home and thus acquire literacy and mathematics in another language than their native language. On more scholastic tasks, native children (speaking Luxembourgish or German) consistently outperform non-native children (not speaking Luxembourgish or German) in German reading comprehension and mathematics in grade 3 (Martin et al., 2015). These differences in

learning performances, however, are evident as early as grade 1 (Hornung et al., 2023 & Hoffmann et al., 2018), and may cause children to be placed in lower educational tracks (Hadjar et al., 2015, 2018).
As most scholastic tests are presented in a specific target language, lower scores might be due to difficulties in understanding the test language rather than limited learning potential (e.g., Greisen et al., 2021; Hickendorf, 2013; Paetsch et al., 2016). In terms of differential diagnostics or diagnostics in general, a test potentially disadvantaging a specific group of test takers runs the risk that the practitioner oversees a necessary field of action. Typical fields of cognitive assessment in the context of education are, e.g., for example, to assess the cognitive abilities of a child as far as possible independently of school content; or to compare whether the child's performance corresponds to its cognitive potential, is significantly below it (so-called underachievers) or significantly above it (so-called overachievers); or to recognize intellectual giftedness as in high cognitive potential or the opposite intellectual disability as in low cognitive potential.
The possible applications are very diverse, and the same test is not suitable for all purposes. In the context of this project, a test battery for grade 3 was developed as these are typically diagnosed after 2 years of formal education to support the diagnostic process of learning disorders more specifically.

Development of FLUX (Fluid Intelligence Luxembourg)

We developed the tablet-based cognitive test battery FLUX (Fluid Intelligence Luxembourg) tailored to the specific needs of multicultural and multilingual educational settings. The assessment is designed to measure the cognitive abilities of grade 3 (Cycle 3.1) children in a fair manner, regardless of their cultural and language background. What makes the test culturally and linguistically fair is that it uses abstract content equally familiar to all children, with language-fair animated instructions. Feedback is given during training items to ensure task comprehension.

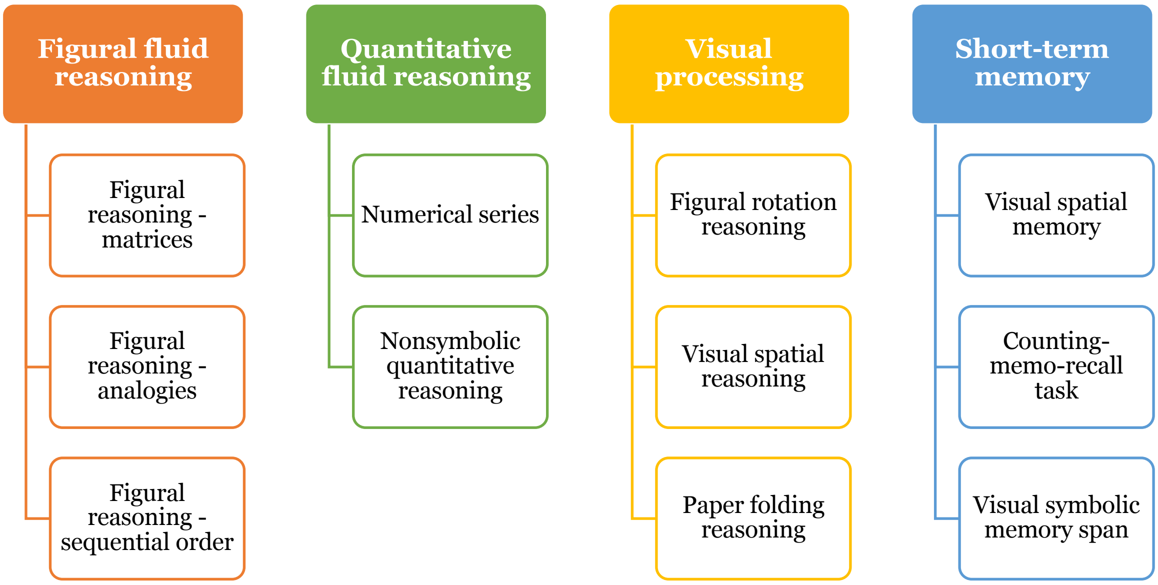
To design the test battery, we referred to theoretical models (mainly the Cattell-Horn-Carroll model of intelligence, McGrew, 2005), recommendations for nonverbal testing (Coleman et al., 1993; Flanagan et al., 2013; Jensen 1980; McCallum, 2017; McGrew & Flanagan 1998), and existing test batteries (e.g., UNIT: Bracken & McCallum, 1998; SON-R 2-8: Tellegen & Laros, 2018; Progressive Matrices: Raven et al., 1998; WISC-V: Wechsler, 2014; CTONI-2: Hammill et al., 2009) as a guide.
After constructing test items based on theoretical models, the consultation of already published but language/culture-specific test procedures, and considerations regarding practicability, these were checked for their psychometric quality using data of approximately 200 multicultural and multilingual third-grade children as part of a pretest.
This pretest was designed to evaluate children's comprehension of instructions and to pilot a pool of more than 450 items to assess the characteristics of each item using item response theory (IRT). This helped to evaluate, for instance, the item discrimination and difficulty. Additionally, item differential functioning (DIF) was applied to ensure fairness in the test by testing whether the test items work comparably in different subgroups. Based on these analyses, some items were removed, some were modified, and some were kept for the main test. Time limits were not implemented because power tests are considered to be more culturally fair than speed tests (Jensen 1980, Kim & Zabelina 2015, Coleman

et al. 1993). Additionally, time predictors are found not to be the best predictors of intellectual performance in children under the age of 10 (Lavergne & Vigneau, 1996).

Following the data analysis, the test material was adapted if necessary and a solid collection of items compiled for nationwide standardization using a representative sample of 703 multicultural and multilingual children.
In addition, within the phase of item development, close attention to an inclusive approach was paid – colorful task contents were chosen to be child-friendly but not overloading to avoid distraction and color combinations were selected for accessibility to colorblind children. This was done in close cooperation with experts in vision, who reviewed all the test material.

The complete FLUX battery includes four domains (*figural fluid reasoning, quantitative fluid reasoning, visual processing, and short-term memory*), each measured by 2 to 3 subtest(s) (see Figure 1).

Figure 1: **Overview of 11 subtests of the language-fair test battery FLUX**



Exemplary detailed presentation of a subtest

In this chapter, we will present the results of one of the 11 subtests, *sequential order,* to illustrate the procedure and present the results per language group of the complete test battery.
In Figure 2, an example of the instruction and training procedure for the subtest *sequential order*, which measures inductive reasoning, is shown. For *the sequential order* task, the children need to complete the sequences of figures accurately.

The fox demonstrates the task to the children using body language (such as smiling and pointing) and language-fair speech sounds (like mumbling, pondering, and expressing an idea). During the explanation, the fox intentionally selects a wrong answer to show the feedback children would receive (a red smiley with the corners of the mouth down) if they chose the wrong answer during training. Then, the correct answer is justified and selected (a green smiley with the corners of the mouth up appears). During the training phase, children are given feedback in the form of a red or a green smiley, which indicates whether they have replied correctly or not. The only selectable option is the green smiley, so they can only proceed if they solved the task correctly by pressing on the smiley. The tasks are designed to be intuitive, and with the help of feedback and instructions, children can understand them easily. However, if they have any doubts, they can ask questions to clarify.

Once they have understood the task, they can move on to the test phase by pressing on a play button. During this phase they will not be allowed to ask any questions nor receive feedback on their solutions.

Figure 2: **Overview of a procedure for the subtest *sequential ordering* to ensure task comprehension in multicultural and multilingual children.**

For *the sequential order* task, the children need to complete the sequences of figures accurately. They have to analyse the sequence of figures presented and find out the rationale (e.g., the figure gets smaller and smaller) to select the correct solution out of four possible options to complete the sequence (see Figure 2, instruction procedure). Test quality criteria, such as the reliability of the task, have been tested (e.g., *McDonald's omega* = .72, indicating good reliability for measuring the same construct *inductive reasoning* across all items).

Impact of language background

To find out whether the task does not disadvantage non-native children, we analysed data from 693 third-grade children from across all school districts in Luxembourg who completed the subtest (M = 8.85 years; SD = 0.66). We collected demographic information from children by administering a questionnaire that included questions about their age and home language(s). Based on the language spoken at home, children were divided into two groups: the non-native language group (N = 417) consisted of children who do not speak Luxembourgish or German at home with at least one parent or caretaker, while the native language group (N = 268) consisted of children who speak Luxembourgish or German at home with at least one parent or caretaker. Results reveal that non-native children (M = 8.05; SD = 2.829) do not significantly differ in performance compared to native children (M = 8.25; SD = 2.786) in this task (F (1,683) = 0.862, p = .353). Consequently, and taken together with the results of statistical analyses at the test level or the items (IRT and DIF), this indicates that the subtest, *sequential order,* can be considered language-fair.

Conclusions

The FLUX project aimed to create a standardized language-fair tablet-based cognitive test battery for multicultural and multilingual third-grade children while applying language-fair instructions alongside non-verbal content. Exemplarily presented results from the *sequential order* task, which is part of the battery, show that both non-native and native children performed equally well. This indicates that it is possible to assess cognitive abilities fairly, regardless of a child's language background.

Furthermore, the cognitive test battery can help practitioners gain detailed insights into a child's cognitive abilities across three domains, allowing them to identify areas of strength and weakness (see Figure 1). This identification is also interesting while applying the test in the framework of a differential diagnosis of a specific learning disorder. Besides specific literacy or mathematical tasks, the language-fair battery can be useful in the diagnostic process to exclude other neurocognitive reasons that might affect learning processes.

Many educational tests are given in a specific language, which can cause some children to underperform due to language barriers instead of a lack of learning ability, i.e., children do not perform up to their potential. FLUX, which uses language-fair content and instructions, could be a valuable tool in identifying these children and providing them with the support they need to succeed.

Overall, the availability of a language-fair test of cognitive abilities will prevent practitioners from relying on tests from other countries that have not been adapted to the Luxembourgish setting in a standardized manner. With FLUX, we have created a psychometrical test that has not only been developed and analyzed according to all the latest standards, but also closes a major gap to prevent the risk of individual students being systematically disadvantaged. This will create a fairer setting by avoiding disadvantaging non-native children and producing invalid test results, as FLUX will allow practitioners to assess cognitive abilities independent of the language background.

References

Bracken, B.A., & McCallum, R.S. (1998). Universal Nonverbal Intelligence Test. Itasca, IL : Riverside.

Coleman, M., Scribner, A. P., Johnsen, S., & Evans, M. K. (1993). A Comparison between the Wechsler Adult Intelligence Scale-Revised and the Test of Nonverbal Intelligence-2 with Mexican-American Secondary Students. Journal of Psychoeducational Assessment 11(3), 250-258.

Flanagan, D. P., Ortiz, S., & Alfonso, V. (2013). Essentials of cross-battery assessment (3rd ed.). New York: John Wiley & Sons.

Greisen, M., Georges, C., Hornung, C., Sonnleitner, P., & Schiltz, C. (2021). Learning mathematics with shackles: How lower reading comprehension in the language of mathematics instruction accounts for lower mathematics achievement in speakers of different home languages. Acta Psychologica, 221, 103456.

Hammill, D. D., Pearson, N. A., & Weiderholt, J. L. (2009). Comprehensive test of nonverbal intelligence (2nd ed.). Austin: PRO-ED.

Hadjar, A., Fischbach, A., Martin, R., & Backes, S. (2015). Bildungsungleichheiten im luxemburgischen Bildungssystem. In LUCET & SCRIPT (Eds.), Nationaler Bildungsbericht Luxemburg 2015 (pp. 34–56). Luxembourg: LUCET & MENJE.

Hadjar, A., Fischbach, A., Martin, R., & Backes, S. (2018). Bildungsungleichheiten im luxemburgischen Sekundarschulsystem aus zeitlicher Perspektive. In LUCET & SCRIPT (Eds.), Nationaler Bildungsbericht Luxemburg 2018 (pp. 58–83). Luxembourg: LUCET & MENJE.

Hoffmann, D., Caroline, H., Sylvie, G., Pascale, E., Ulrich, K., Fischbach, A., (2018). Schulische Kompetenzen von Erstklässlern und ihre Entwicklung nach zwei Jahren. In LUCET & SCRIPT (Eds.), Nationaler Bildungsbericht Luxemburg 2018 (pp. 84–95). Luxembourg: LUCET & MENJE.

Hornung, C., Kaufmann, L. M., Ottenbacher, M., Weth, C., Wollschläger, R., Ugen, S., & Fischbach, A. (2023). Early childhood education and care in Luxembourg. Attendance and associations with early learning performance. Esch-sur-Alzette, Luxembourg: Luxembourg Center of Educational Testing (LUCET). doi:10.48746/epstanalpha2023pr

Jensen, A. (1980). Bias in mental testing. New York: Free Press.

Martin, R., Ugen, S., & Fischbach, A. (Eds.). (2015). Épreuves Standardisées: Bildungsmonitoring für Luxemburg. Nationaler Bericht 2011 bis 2013. Esch/Alzette : University of Luxembourg, LUCET.

McCallum, R. S. (2017). Handbook of Nonverbal Assessment. Springer.

McGrew, K. S., & Flanagan, D. P. (1998). The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment. Boston: Allyn & Bacon.

McGrew, K.S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. Past, present and future. In D.P. Flanagan, & P.L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (2nd ed., pp. 136–181). New York: Guilfor

Paetsch, J., Radmann, S., Felbrich, A., Lehmann, R., & Stanat, P. (2016). Sprachkompetenz als Prädiktor mathematischer Kompetenzentwicklung von Kindern deutscher und nicht-deutscher Familiensprache. Zeitschrift Fur Entwicklungspsychologie Und Pädagogische Psychologie, 48(1), 27–41.

Raven, J., Raven, J. C., & Court, J.H. (1998). Manual for Raven's Progressive Matrices and Vocabulary Scales. Oxford, UK: Oxford Psychologists Press.

Schaap, P. (2011). The differential item functioning and structural equivalence of a nonverbal cognitive ability test for five language groups. SA Journal of Industrial Psychology, 37(1). https://doi.org/10.4102/sajip.v37i1.881

Tellegen, P.J., Laros, J.A., & Petermann, F. (2018). SON-R 2-8: Non-Verbaler Intelligenztest. I. Technisches Manual; Göttingen, Germany: Hogrefe.

Ugen, S., Schiltz, C., Fischbach, A., & Pit-ten Cate, I. M. (2021). Lernstörungen im multilingualen Kontext – Eine Herausforderung. Lernstörungen im multilingualen Kontext. Diagnose und Hilfestellungen. Melusina Press.

Wechsler D. (2014). Wechsler intelligence scale for children (5th ed.). Bloomington, MN: NCS Pearson.