Measurement of gains in intelligence – an example of a study design and a method of data analysis

Anna Hawrot*, Aleksandra Jasińska*

Educational Research Institute, Warsaw, Poland
a.hawrot@ibe.edu.pl, a.jasinska@ibe.edu.pl

Introduction

In order to estimate absolute differences between two measurements of any variable, it is necessary that they are **expressed on a common scale**. The easiest way to achieve that is to use the same test in subsequent measurements. However, sometimes it is not possible, i.a., due to the fact that the test measures too limited range of the variable of interest.

Such a situation may happen, e.g., when level of the measured variable changes as a result of developmental processes. Usually, the problem is solved through the construction of separate age-adjusted versions of a test (i.e. Wechsler Intelligence Scale for Children and Wechsler Adult Intelligence Scale). Nevertheless, sometimes even age-adjusted tool may not be appropriate and ceiling or floor effect might be observed.

However, using two different tests in subsequent measurements results in two incomparable scores that do not allow to estimate the absolute difference in the variable level between time points. Thus, if a test appropriate in both measurements is not available, other ways to achieve the goal must be considered.

Research problem

Situation similar to the aforementioned one arose in Educational Value-Added research project run in Poland. One of side issues of the study was to examine the **absolute mean gains in students' intelligence** during three years of schooling.

As **cognitive functioning gradually improves over childhood, puberty and early adulthood** (e.g., McArdle, Ferrer-Caja, Hamagami & Woodwock, 2002), gains in intelligence in course of lower secondary school (LSS) were expected. The pace (scale) of these changes was inquired.

Unfortunately, left-sekewd distribution of **Raven's Standard Progressive Matrices** (RSPM) test row scores in the **first measurement** suggested that this test would be too easy for participants in the second measurement. Thus, it was decided to administer subsets of items of two intelligence tests: **RSPM and Raven's Advanced Progressive Matrices** in the **second time point**.

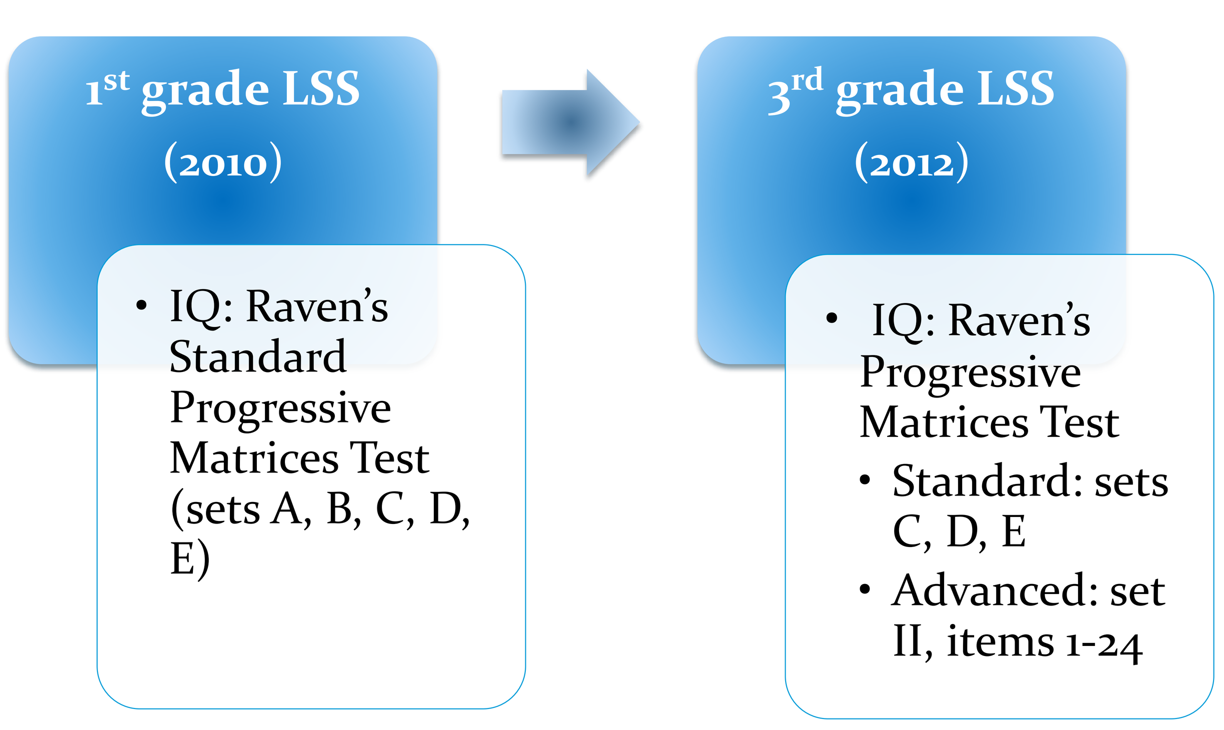
Method

Data. Data gathered during **Educational Value-Added research project** (www.ewd.edu.pl) were used.

Participants

5223 lower secondary school (LSS) students
291 classes, 150 schools

Design and measures



Test equating

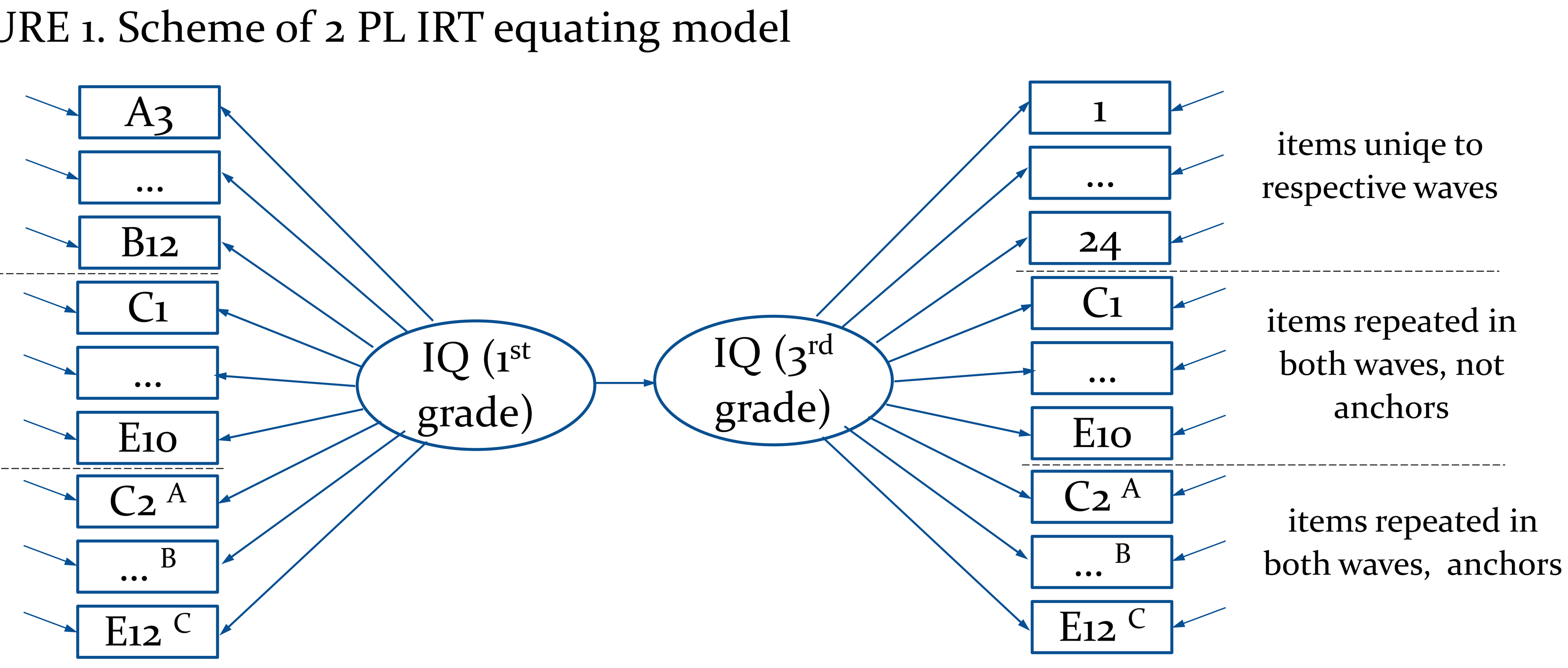
The results of two measurements were equated with use of **the nonequivalent groups with anchor test design (NEAT) with internal anchor**. The statistical model was modified to take account of administering the tests to the same sample of students. **Differential Item Functioning (DIF)** of the anchor items was verified; items with substatial DIF were excluded from the anchor set. In total, 30 items were used as anchor items. See Table 1 for details.

TABLE 1. Structure of IQ tests used in both time points

RPM Test - version	Item	1st grade	3rd grade	Anchor items
Standard	A3-B12	+		
	C1	+	+	
	C2-C10	+	+	+
	C11-C12	+	+	
	D1-E6	+	+	+
	E7	+	+	
	E8	+	+	+
	E9-E10	+	+	
	E11-E12	+	+	+
Advanced	1-24		+	

The **scheme of the equating model is presented** in Figure 1. Parameters (difficulty and discrimination) of anchor items were constrained equal across measurements, whereas parameters of non-anchor items were estimated freely. Mean and standard deviation were fixed respectively to 0 and 1 in the 1st measurement and estimated freely in the 2nd measurement. Clustering was taken into account. WLSMV estimator was used. Such a model is equivalent to 2 PL IRT model (Bartholomew, Knott & Moustaki. 2011). Analyses were performed with Mplus 6.11 software. **The model fitted data adequately** (RMSEA = 0.011, TLI = 0.954).

FIGURE 1. Scheme of 2 PL IRT equating model

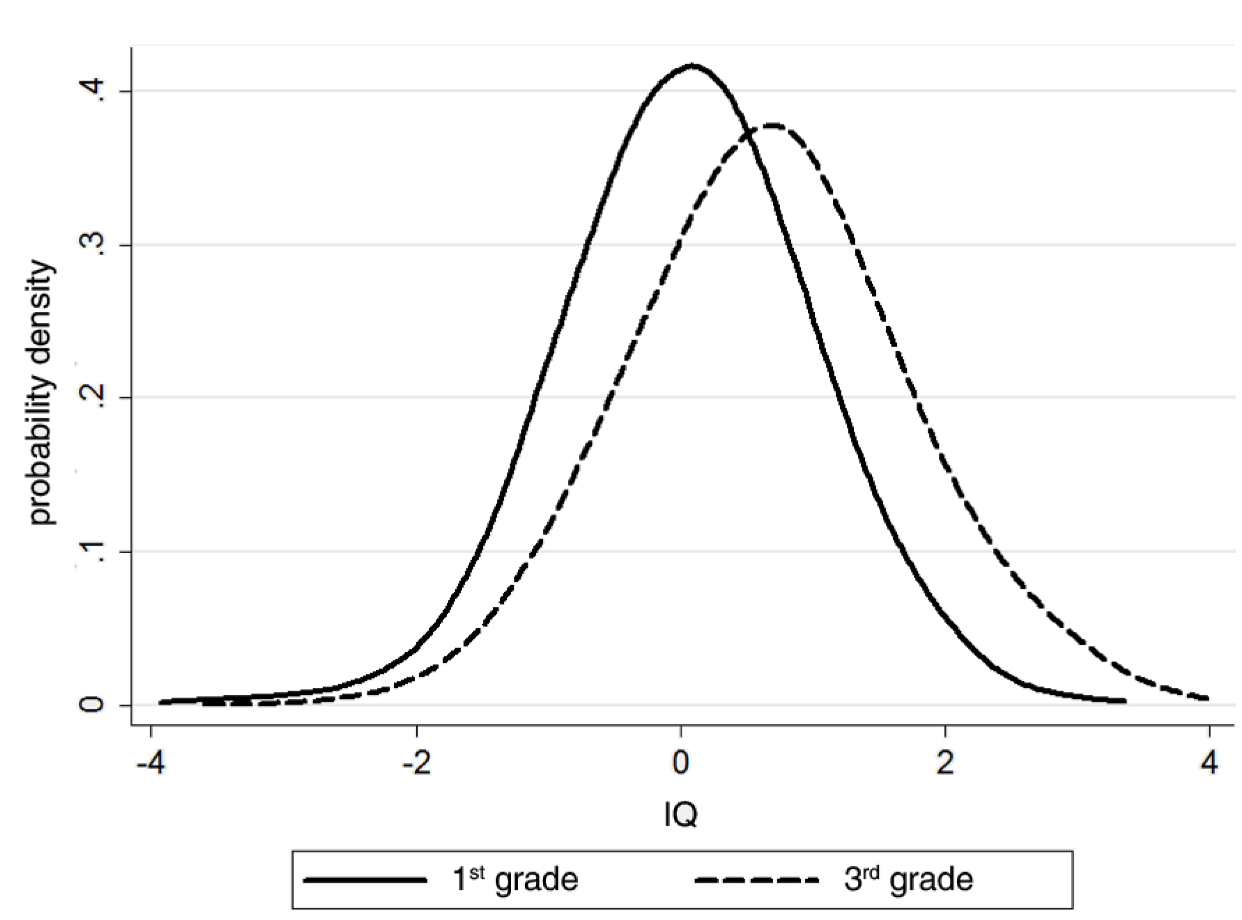


Letter superscripts in variables' names indicate parameter constraints.

Results

The model allowed us to equate the results of two intelligence measurements (express them on a common scale) and to estimate intelligence gains between the first and the third grade. Distributions of IQ are presented in Figure 2.

FIGURE 2. IQ distributions in grade 1 and grade 3



The results of the 1st and the 3rd grade measurements were **strongly correlated** (latent correlation equaled **0.71**).

Students' intelligence rose on the average by **0.6 standard deviation** of the 1st grade results.

References: ask the authors.

HUMAN CAPITAL NATIONAL COHESION STRATEGY · IBE · EDUCATIONAL VALUE ADDED · EUROPEAN UNION EUROPEAN SOCIAL FUND