Contents lists available at ScienceDirect

Intelligence

An item-level examination of the Flynn effect on the National Intelligence Test in Estonia

William Shiu [a], A. Alexander Beaujean [a,*], Olev Must [b], Jan te Nijenhuis [c], Aasa Must [d]

[a] Baylor University, United States
[b] Tartu University, Estonia
[c] University of Amsterdam, The Netherlands
[d] Estonian Defence College, Tartu, Estonia

ARTICLE INFO

Article history:
Received 9 December 2012
Received in revised form 6 May 2013
Accepted 26 May 2013
Available online 20 June 2013

Keywords:
Flynn effect
Item response theory
Estonia
Invariance

ABSTRACT

This study examined the Flynn effect (FE; i.e., the rise in IQ scores over time) in Estonia using the Estonian version of the National Intelligence Tests (NIT; Haggerty, Terman, Thorndike, Whipple & Yerkes, 1919; National Research Council, 1920). Using secondary data from two cohorts (1934, $n = 890$ and 2006, $n = 913$) of students, we analyzed the NIT's subtests using item response theory (IRT). For each subtest, we first examined invariance in all the items and then linked the latent variable (θ) scores between the two cohorts using the invariant items. The results showed that there was a FE in θ for all subtests except one, although there was much variability in the FE magnitude, ranging from an effect size of 0.24 (3.60 IQ points) to 1.05 (15.75 IQ points). In addition, this study showed there was a decrease in the variability of θ for all the subtests, although only two of the subtests showed large decreases (approximately .50 standard deviations). Last, the subtests' precision of measuring θ was very similar at both time points.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

The Flynn effect (FE) is the rise in IQ scores over time (approximately 3 IQ points per decade or .3 points per year; Flynn, 2007; Neisser, 1998). The FE has been found on every inhabitable continent (e.g., Flynn, 1987; Flynn & Rossi-Casé, 2012; Pietschnig, Voracek, & Formann, 2010; te Nijenhuis, Cho, Murphy, & Lee, 2012; te Nijenhuis, Murphy, & van Eeden, 2011), and across a wide range of abilities (e.g., Howell, 2008; Kanaya, Scullin, & Ceci, 2003; Sanborn, Truscott, Phelps, & McDougal, 2003; Wai & Putallaz, 2012). Moreover, due to its prominence, it is now a part of many legislative discussions that concern cognitive ability assessment (e.g., Ceci, Scullin, & Kanaya, 2003; Flynn, 2006; Kanaya & Ceci, 2007; Young, Boccaccini, Conroy, & Lawson, 2007).

Although many causal theories have been put forth, the cause of the FE remains inconclusive. Some think the FE represents a genuine rise in cognitive ability due to, e.g., better nutrition (Cohen, Flament, Dubos, & Basquin, 1999; Lynn, 2009; Sigman & Whaley, 1998), increased cognitive stimulation (Blair, Gamsonb, Thornec, & Bakerd, 2005; Teasdale & Owen, 1987), or change in family structure and fertility patterns (Mingroni, 2007). Other researchers argue that the FE could just as easily be due to changes in the test as it is due to changes in the individuals taking the test (i.e., psychometric artifact; Beaujean & Osterlind, 2008; Brand, 1987), and argue that more psychometric work in the FE needs to be done before any strong causal theories should be developed (McGrew, 2010; Rodgers, 1998).

1.1. Measuring the Flynn effect

FE studies typically involve one of two types of designs. The first design is to examine scores from two or more versions of the same instrument (or two different instruments

* Corresponding author at: Department of Educational Psychology, One Bear Place #97301, Waco, TX 76798-7301, United States. Tel.: +1 254 710 1548; fax: +1 254 710 3265.
E-mail address: Alex_Beaujean@Baylor.edu (A.A. Beaujean).

0160-2896/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.intell.2013.05.007

normed at different times) administered to a single sample at a single time point (e.g., Covin, 1977). The second design to assess the FE is to compare data from a single instrument administered to two or more (ostensibly) similar samples from different generations (e.g., Flynn, 1987). For either design, the most common method used to measure the FE is to compare mean differences in aggregated scores (e.g., Full scale IQ, Verbal IQ). In doing so, investigators make an implicit assumption that the test scores are measuring the sameconstruct(s) the same way (i.e., invariance; Meredith & Teresi, 2006; Millsap, 2011).

1.2. Measurement invariance

The validity of between-group test score comparisons is threatened if items operate differently among groups (Kane, 2006; Messick, 1989; Ployhart & Vandenberg, 2010). If one cannot be sure that an instrument is measuring the same construct the same way at both time periods, then one cannot be sure that any observed group difference of the instrument's score are due to measuring the constructs differently or a true difference between the groups (Little, 1997; Steenkamp & Baumgartner, 1998; Thompson & Green, 2006; Vandenberg & Lance, 2000). Beaujean and Sheng (2013) make the following analogy comparing means from non-invariant test scores is akin to comparing average temperatures at two different geographic locations with thermometers that use different scales. While mean differences could be due to different temperatures, they could also be the result of the scales having different origins (e.g., Fahrenheit vs. Rankine), different units (e.g., Kelvin vs. Rankine), or both (e.g., Fahrenheit vs. Kelvin). Consequently, before comparing scores between groups, it is important to assess that the instruments are measuring the same construct, the same way (Millsap, 2011; Yoo, 2002).

1.2.1. Investigating measurement invariance

Typically invariance is examined using multi-group confirmatory factor analysis (MG-CFA) (Horn & McArdle, 1992). CFA is a very general latent variable framework that can handle both continuous indicators (traditional factor analysis) and categorical indicators (sometimes called binary factor analysis or item response theory [IRT]) (Bartholomew, Knott, & Moustaki, 2011). If there is at least strong invariance on the instrument among the groups, then group differences on the instrument's scores are due to group differences in the latent constructs the instrument is measuring and not a result of measurement artifacts or cultural differences. For the latent variables to be comparable among groups, at least three conditions must exist (Cheung & Rensvold, 2002; Little, 1997):

1. The indicators for the latent variable have the same configuration among the groups; that is, the groups should have the same number of latent variables, the same number of indicators, and the same pattern of fixed and free parameters (*configural invariance*).
2. The relationships between factors and indicators (i.e., loadings/pattern coefficients) are the same among the groups, thus establishing equivalence of the metrics of the latent variable(s) among groups (*weak/scalar invariance*).
3. The indicators' intercepts are the same among the groups, thus establishing equivalence of the latent variable's origin (*strong/scalar invariance*). In some situations, there is invariance in some indicators' loadings and intercepts, but not all of them, a situation Byrne, Shavelson, and Muthen (1989) called *partial invariance*. Although much more work needs to be done in this area (Vandenberg, 2002), the basic premise behind partial invariance is that as long as there is configural invariance and there are “enough” invariant indicators, the latent variable(s) can still be compared across groups. In such situations, the measures can be considered to be alternate forms: measuring the same latent variable, but using different indicators (although in situations involving the FE, there will be item overlap between the forms). To be able to compare the scores from the alternate forms, however, they must be first be equated.

1.2.2. Equating

The purpose of equating is to convert item and ability estimates from different measurement instruments (or alternate versions of the same instrument given to different populations) to a common scale to be able to compare the examinees' abilities (Baker, 1984; Dorans, 2004). There are two types of equating: horizontal and vertical. Horizontal equating is typically used to equate scores on alternate forms of a test given to equivalent groups, while vertical equating is typically used to equate scores on tests that differ in (overall) difficulty that are given to groups that differ in amount of the trait the test is measuring. For vertical equating, typically a common set of items (i.e., anchor items) is used across at least two versions of the test, which are used to determine the link needed to place the test scores on the same scale (Baker, 1984). With horizontal equating, the equivalence of the examinees allows for the alternate forms to be linked, although the equating process is greatly strengthened when there are anchor items used in this form of equating as well (Kolen & Brennan, 2004).

Such methods can readily be transferred to investigating the FE. If, as some hypothesize, subsequent generations are increasing in cognitive ability, then vertical equating can be used to transform the ability scales from different instruments for the different groups. On the other hand, if just the test properties are changing over time, then horizontal equating methods can be used to equate the scales from different measures for the equivalent groups.

1.3. Estonia

Estonia is a small country located in north-eastern Europe. The country covers approximately 45,200 km^2 and has a population of approximately 1.32 million. For many centuries, Estonia was the border between the western and the eastern world. The Estonian language belongs to the Finno-Ugrian branch of the Uralic language family, and many Estonians consider themselves to be a member of Nordic nations. Estonians declared their political independence in 1918, but the country was occupied by the Soviet Union in 1939/1940. After the collapse of Soviet regime in 1991, Estonia re-established its independence.

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات