Learning and Individual Differences 45 (2016) 128–132

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Another look at the Spearman's hypothesis and relationship between Digit Span and General Mental Ability

Yoav Ganzach

Tel Aviv University, Israel

ARTICLE INFO

Article history:
Received 23 February 2015
Received in revised form 25 August 2015
Accepted 28 December 2015

Keywords:
Intelligence
Spearman's hypothesis
Digit Span
Blacks
Hispanics

ABSTRACT

In a classic paper in intelligence research, Jensen and Figueroa (1975) examined differences between Digit Span Forward (DSF) and Digit Span Backward (DSB), both between groups and between individuals within groups. This paper was the basis for the later development of the Spearman's hypothesis, which states that the stronger the association between General Mental Ability (GMA) and a test, the larger the between-groups differences in test scores. In the current study we re-examine Jensen and Figueroa's results on the basis of a large, nationally representative database. One of our results replicates Jensen and Figueroa earlier results: Consistent with Spearman's hypothesis, we find that the difference between Blacks and White is larger in DSB than in DSF. However, in contrast to the Spearman's hypothesis, we find that the Hispanic–White difference is larger in DSF than in DSB. In addition, in contrast to Jensen and Figueroa (1975), we find that within groups proxies of GMA tend to be more strongly associated with DSF than DSB.

1. Introduction

Understanding between-groups and between-individuals variations in specific mental abilities is important because it may shed light on the origin of both between groups and between individual's variations in GMA, or General Mental Ability (g). Particularly important in the development of the research in this area (see Jensen, 1998, pp. 369–372, for an historical overview), was a study by Jensen and Figueroa (1975)(to be labeled JF) which examined Digit Span Forward (DSF), a test in which subjects are asked to recall a list of digits according to the order in which they were recited, and Digit Span Backward (DSB), a test in which subjects are asked to recall the list backward. Jensen and Figueroa (1975) formulated and confirmed two major hypotheses, one concerned a within groups difference between DSF and DSB and the other concerned a between groups difference between the two. First, they argued that, within groups, the relationship between GMA and DSB is stronger than the relationship between GMA and DSF (hypothesis 1 in JF paper). And second, they argued that DSF showed a small Black–White (B–W) difference in test scores, while DSB showed a large difference: Whites recall more digits than Blacks in this test (hypothesis 2 in JF paper).

Jensen and Figueroa (1975) explanation for the within-groups difference between DSF and DSB was that DSB is a more mentally demanding task than DSF, and therefore more strongly associated with GMA. Note that in JF study, the hypotheses about the association between the two Digit Span tests were derived from an a priori evaluation of the mental demands of the two tests (see JF, pp. 882–883). In this respect, JF study is different from later work which did not attempt to a priori evaluate tests' mental demands and implicitly assumed that the association between GMA and a specific test is a measure for the mental demand of a test (but see Jensen, 1993). As only very few studies that attempted to replicate JF result about the association between tests' mental demands and GMA in general and between the two Digit Span tests and GMA in particular, and in view of the importance of this result in our current understanding of between-groups variations in mental tests scores (see below), the first purpose of the current study is to re-examine JF finding regarding the within group association between the two Digit Span tasks and GMA (JF hypothesis 1).

Turning to JF hypothesis 2, the explanation offered by Jensen and Figueroa to the small B–W difference in DSF vs. the large difference in DSB was similar to the explanation they offered to the within group association between these two tests and GMA: Since DSB is more strongly associated with GMA than DSF, it more strongly reflects the basic differences between Blacks and Whites in GMA. Later, Jensen and others extended this idea and argued that B–W differences in a variety of other mental tests are related

E-mail address: yoavgn@post.tau.ac.il.

http://dx.doi.org/10.1016/j.lindif.2015.12.026
1041-6080/

to the strength of the association between GMA and these tests, an idea that was labeled 'the Spearman's hypothesis'.[1] They also extended the methodology by which this association was assessed, estimating the association between mental tests and GMA, not only based on an a-priori evaluation of this association, but also on the correlation between test g-loading – its loading on the first principle component extracted from a battery of mental tests – and group differences in the test (i.e., the method of correlated vectors). This methodology, however, was criticized by several authors (e.g., Ashton & Lee, 2005; Guttman, 1992; Gustafsson, 1992; Dolan & Lubke, 2001). In this sense, studies that examine between group differences relying on a-priori evaluation of the association between the mental test and GMA are immune to these criticisms. Furthermore, note that such studies can directly demonstrate that between groups differences are not due to cultural differences. In particular the B–W difference between DSF and DSB cannot be easily explained by cultural differences between the two groups, since both tests have the same content. However, despite these advantages of examining the Spearman's hypothesis by comparing B–W difference in the two Digit Span tasks, and despite the fact that Jensen and Figueroa's (1975) finding regarding this difference was critical to the development of Spearman's hypothesis, and was central for the debate about the sources of between-groups differences in intelligence, there is, to the best of our knowledge, only one published replication of these findings (Jensen & Osborne, 1979). Thus, a second purpose of the current study is to attempt to replicate this finding (associated with JF hypothesis 2) in a large representative sample.[2]

A third purpose of the current study is to examine the generality of the Spearman's hypothesis on the basis of differences between DSF and DSB not only for Blacks and Whites, but also for Hispanics. Previous studies of the Spearman's hypothesis involving comparisons other than Black–Whites comparisons did not yield consistent results. Some yielded clear support for this hypothesis (e.g., te Nijenhuis, David, Metzen & Armstrong, 2014; Rushton, 2002; Rushton, Skuy, & Fridjohn, 2002; Sandoval, 1982; and many studies by te Nijenhuis and his co-authors such as te Nijenhuis, Al-Shahomee, van den Hoek, Grigoriev, & Repko, 2015a,b; te Nijenhuis, van den Hoek & Armstrong, 2015; te Nijenhuis & van der Flier, 1997; te Nijenhuis & van der Flier, 2005), while others (e.g., Helms-Lorenz, Van de Vijver, & Poortinga, 2003; Lynn & Owen, 1994; Nagoshi, Johnson, DeFries, Wilson, & Vandenberg, 1984; Rushton, Skuy, & Fridjohn, 2003) failed to find such support. To the best of our knowledge, only one study (Hartmann, Hye, & Nyborg, 2007), directly examined a Hispanic group, obtaining conflicting results. All these studies, however, involved relatively small non-representative samples.

2. Method

2.1. Participants and procedure

The data was taken from an ongoing longitudinal study, the NLSY79 Children survey (the NLSY79C). The NLSY79C is based on an earlier survey, the 1979 cohort of the National Longitudinal Study of Youth (the NLSY79), which is a probability sample of 12,686 Americans (with an oversampling of Afro-Americans, Hispanics and economically disadvantaged whites) born between 1957 and 1964. The NLSY79C sample frame consists of all children born to female NLSY79 respondents. The number of children who were interviewed increased from 5255 in 1986, the initial child collection survey year, to a total of 11,504 in 2010, the last survey year we use in the current paper. The children were surveyed every two years, and at each survey they received a series of cognitive tests. In addition, the mothers were also interviewed, and information about the household was collected by the interviewer.

The interviews of the children were typically conducted in the home of the child's mother by experienced, specially trained field staff. Child interviews through 1992 were conducted primarily in person using paper and pencil. Beginning in 1994, the interviews were administered using Computer-Assisted Personal Interviewing (CAPI). Spanish translations of several of the test instruments have been made available to respondents with limited proficiency in English. However, the number of children who were assessed in Spanish was very low. For example, in 2000, the number of children assessed in Spanish was fewer than 10, and at this year, most of the Spanish language parents would have resided in the U.S. for more than two decades.

2.2. Variables and measurement

2.2.1. Digit Span Forward (DSF) and Digit Span Backward (DSB)

In the first part of the test the child listens and repeats a sequence of numbers said by the interviewer. In the second part, the child listens to a sequence of numbers and repeats them in reverse order. In both parts, the length of each sequence of numbers increases as the child correctly responds. The test was given to children between the ages 7–12. In the analysis we used both the raw scores of each of the two tests (i.e., the actual number of digits remembered correctly) as well as the scaled scores, each scaled to a mean of 10 and standard deviation of 3 within each 4-month age interval.

2.2.2. GMA

Unfortunately, the NLSY79C does not have direct measures for GMA. We therefore used the following two measures as proxies for GMA: The child score on the Peabody Picture Vocabulary Test (PPVT) and the GMA of the child's mother.

2.2.3. The Peabody Picture Vocabulary Test (PPVT)

This test "measures an individual's receptive (hearing) vocabulary for Standard American English and provides, at the same time, a quick estimate of verbal ability or scholastic aptitude" (Dunn & Dunn, 1981). The test is considered a reasonable estimate for GMA (e.g., Campbell, Bell, & Keith, 2001; Childers & Durham, 1994; Bell, Lassiter, Matthews, & Hutchinson, 2001; Snitz, Bieliauskas, Crossland, Basso, & Roper, 2000; Smith, Smith, & Dobbs, 1991. But see Bracken & Prasse, 1982; Hodapp & Gerken, 1999). We chose this test as a measure for children's GMA since it provided the best estimation of GMA among the tests that were available in the NLSY79C.[3]

The test consists of 175 vocabulary items of generally increasing difficulty. The child listens to a word uttered by the interviewer and then selects one of four pictures that best describes the word's meaning. A child's entry point into the assessment is based on his or her PPVT age. In a few cases, a Spanish version of the PPVT-R was used until 2000. We used the standard age-normed scores of the PPVT which are provided by the NLSY79C with a mean of 100 and a standard deviation of 15.

[1] Although originally the Spearman's hypothesis referred to the observation that the stronger the association between a mental test and g, the larger the B–W difference in this test, following other authors (e.g., Jensen, 1985; Rushton & Jensen, 2005) in the current paper we refer to the Spearman's hypothesis as referring to a number of related phenomena including the observation that the stronger the association between a mental test and g, the larger the B–W difference in this test; the observation that the stronger the association between a mental test and g, the larger the difference in this test between any two groups that differ in their average g; the observation that the larger the mental demand associated with a test, the larger its association with g; as well as the original two-level theory of mental abilities developed by Jensen and Figueroa suggesting that Blacks and Whites differ in complex cognitive processing, but not in rote learning and memory.

[2] Jensen and Figueroa (1975), had three more hypotheses, all of them related to developmental differences between DSF and DSB. As these hypotheses are not related to Spearman's hypothesis, they are beyond the scope of the current paper. They are examined elsewhere (Ganzach & Gotlibovski, 2014).

[3] The other mental tests that are available in the NLSY79C are the PIAT-math, PIAT-reading recognition, and PIAT-reading comprehension. These are all considered achievement tests.