Intelligence 41 (2013) 439–451

Contents lists available at SciVerse ScienceDirect

Intelligence

journal homepage:

# Spearman's law of diminishing returns: A statistical artifact?

Aja L. Murray [a,b,*], Hayley Dixon [a], Wendy Johnson [a,b,c]

[a] Department of Psychology, University of Edinburgh, UK
[b] Centre for Cognitive Ageing and Cognitive Epidemiology, UK
[c] Department of Psychology, University of Minnesota–Twin Cities, USA

ARTICLE INFO

*Article history:*
Received 22 February 2013
Received in revised form 1 June 2013
Accepted 7 June 2013
Available online xxxx

*Keywords:*
Spearman's Law of Diminishing Returns
SLODR
Differentiation
Ability structure
Skew

ABSTRACT

Spearman's Law of Diminishing Returns (SLODR) is the idea that the structure of human cognitive ability is more differentiated and *g* a weaker determinant of cognitive performance at higher levels of ability. In this study, we distinguish between 'traditional' methods of testing SLODR and 'contemporary' methods of testing SLODR. It is the former set of methods from which the vast majority of the evidence base for SLODR derives. We demonstrated that it is easy to mimic SLODR and reverse SLODR effects in these traditional methods of assessing SLODR by using data with skewed observed variable distributions. The skewness magnitudes did not need to be large to produce these effects and they fell well within the range of values that are usually considered unproblematic for parametric statistic analysis. In simulated datasets, positive subtest skewness resulted in SLODR and negative subtest skewness resulted in reverse SLODR. In contemporary methods of testing SLODR, non-linear *g*-loadings or a skewed *g* are assumed to reflect evidence for SLODR. When we applied contemporary methods of testing SLODR to these data, there was evidence of heteroscedastic residuals but no evidence of non-linear *g*-loadings or skewed *g* distributions. We broadly replicated the effects of subtest skew from these simulated datasets in real data from the Minnesota Study of Twins Reared Apart. Results imply that traditional methods of assessing SLODR cannot distinguish between effects due to subtest characteristics that have nothing to do with differences in ability structure at different levels of *g* and true SLODR effects. This calls into question the empirical support for SLODR.



## 1. Introduction

In 1927, Charles Spearman made an observation that has since come to be known as 'Spearman's Law of Diminishing Returns (SLODR; Spearman, 1927)'. This is the idea that at higher levels of ability, its structure is more differentiated and *g* is a weaker contributor to cognitive performance. Since Spearman proposed this idea, models of intellectual ability have evolved to include several strata of ability factors, superseding the single-*g*-factor theory that prevailed in his time (e.g. Gustafsson, 2001). Empirical investigations of SLODR have lagged behind these developments, with second-order modelling of SLODR being developed only in the last few years (Molenaar, Dolan, & van der Maas, 2011; Molenaar, Dolan, Wicherts & van der Maas, 2010; Reynolds & Keith, 2007; Reynolds, Keith, & Beretvas, 2010). As a result, the vast majority of research into SLODR has used a one-factor model of *g*.

A typical study of this sort (which will henceforth be referred to as the 'traditional method') has involved splitting a sample of individuals assessed on a battery of cognitive ability tests into a high-ability group and a low-ability group (Jensen, 2003). Individuals would be assigned to these groups based on a sum or other composite score (e.g. Hartmann & Teasdale, 2005) or on their scores on a single subtest (the selection subtest), which would then be omitted from subsequent analyses (e.g. Deary et al., 1996). The rationale behind omitting the selection subtest was that this ensured that the measure used to estimate *g* was distinct from those used to test SLODR

* Corresponding author at: Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK.
*E-mail address:* s0785823@staffmail.ed.ac.uk (A.L. Murray).

0160-2896/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.intell.2013.06.007

(Jensen, 2003). Regardless, SLODR was then assessed by taking measures of the inter-relatedness of tests within the high- and low-ability groups. This was done using the average inter-correlations of the subtests or the eigenvalues or percentages of variance explained by the first principal component (PC1) (Jensen, 2003). If the degree of inter-relatedness was larger in the low-ability group than in the high-ability group, this was taken as evidence of SLODR.

There are several issues with this traditional method. First, it necessitates the artificial dichotomisation of the ability continuum at an arbitrary point. This may result in reduction of the statistical power to detect SLODR but also introduces non-normality within ability groups, undermining the parametric assumptions upon which the subsequent comparison of the two groups is based (Kane & Brand, 2006). Second, if a composite score is used to estimate *g*, then this estimate is not independent of the data used to estimate SLODR and results in possible distortions of within-group structure, related to attenuation of subtest inter-correlations (Jensen, 2003). If, on the other hand, a single selection subtest is used to estimate *g* and then omitted from subsequent analyses, then this is likely to be a poor estimate of *g* (Hartmann & Nyborg, 2004) and could, for example, produce different patterns of regression to the mean depending on the specific subtest used. Within-group inter-correlations will, in fact, be affected even if the selection subtest is omitted (Muthén, 1989; Pearson, 1903). Third, the use of a one factor model of *g* is generally considered suboptimal with respect to obtaining reliable estimates of *g* (Major, Johnson & Bouchard, 2011). In the context of testing SLODR it also results in the conflation of a number of possible differentiation effects at different levels of ability structure i.e. in individual subtests, specific ability factors such as verbal or spatial ability, and *g* (Molenaar et al., 2011).

Perhaps the most important and yet least discussed potential limitation of these traditional methods, however, is their inability to separate out the effects of measurement or sampling characteristics of the data – which have nothing to do with ability structure – from true SLODR effects. It is worth considering the effects of data skew, in particular. If SLODR is true, it would tend to cause skew at the subtest and latent level (Molenaar, Dolan, & Verhelst, 2010; Molenaar et al., 2011). However, data skew is an almost universal feature of psychological data that can have a range of other causes which are unrelated to true differences in ability structure (Bishara & Hittner, 2012; Micceri, 1989), thus creating the potential for circularity in the relation between subtest characteristics and SLODR. Skew could, for example, be introduced by various aspects of the measurement properties of subtests or items such as those resulting from the presence of floor or ceiling effects, trait-level dependent discriminability or reliability, the use of sum scores from a small number of items or a large number of disproportionately easy or difficult items (e.g. Der & Deary, 2003; Tucker-Drob, 2009).

Skew can also arise from the sampling characteristics of the data, in particular from inadequate or truncated sampling (Arnold & Beaver, 2002). Truncated sampling has obvious relevance for intellectual ability research because individuals with intellectual disabilities tend not to be included in general psychological research due to ethical and recruitment difficulties (Arscott, Dagnan & Kroese, 1998). Thus, individuals with lower IQs may come to be under-represented in studies of intellectual ability. These truncation effects could also result from the more general sample selection effects widely observed in psychological research which appear to be associated with characteristics that include intellectual ability and correlated traits (see Arellano-Valle, Branco, & Genton, 2006; Madhyastha, Hunt, Deary, Gale, & Dykiert, 2009).Thus, although SLODR would be expected to cause skew, given the myriad possible causes of skew other than SLODR, the tendency for ability subtests to be skewed in both positive and negative directions, and the ubiquity of skewness in psychological data generally, it would be unrealistic to assume that SLODR (or reverse SLODR) is the sole cause of observed subtest skewness.

Many statistical techniques tolerate some violations of strict normality quite well, so that rules of thumb such as acceptability of skewness less than |1| are in common use. This may have led researchers to neglect to consider the implications of small levels of skewness on tests of SLODR. Small skews do, however, matter in some situations, and SLODR may be one such situation. If it is assumed that SLODR is false, but that there is observed variable skew for some other reason, it is possible to predict how this might affect tests of SLODR by considering the effects of skew on the methods used to assess SLODR. The same logic can be used to predict how true SLODR effects could be masked by observed variable skew. In traditional analyses of SLODR, the effect of univariate skew should depend on exactly how the data are analysed and the relative skews of the subtests used to assess SLODR and the measure used to define ability groups. It was considered an important practice in traditional analyses to match groups for variance in the estimate of *g* used to define ability groups (Te Nijenhuis & Hartmann, 2006). This was because subtest variance differences across groups could mask SLODR when there were larger variances in the high-ability group and mimic SLODR when there were larger variances in the low-ability group (Hartmann & Teasdale, 2004). If, however, a selection subtest used to define these ability groups was skewed, variance may have been matched for this measure but markedly different in the remaining subtests, causing the very effects which matching for variance in the subtest was designed to prevent. When a positively skewed selection subtest is used to select ability groups matched for variance, the low-ability group will likely have larger variances for most subtests. This is because, in a positively skewed distribution, observations below the mean are more clustered together and observations above the mean are more scattered. Thus, equating variances in low- and high-ability groups results in setting the threshold for group membership above the mean to compensate for these differences in scatter. If the remaining subtests are approximately normally distributed, there is likely to be more variance in these remaining subtests in the low-ability group simply because more participants have been assigned to that group. The opposite is true if a negatively skewed selection subtest is used. Thus, use of positively skewed selection subtests may have produced illusory SLODR and negatively skewed selection subtests may have produced illusory reverse SLODR in previous studies. Indeed, in previous studies that have used multiple selection subtests and reported their skew, it is possible to observe associations between selection subtest skew and SLODR effects (Abad, Colom,