Will the Real Discrepant Learning Disability Please Stand Up?

Wim Van den Broeck

Abstract

Willson and Reynolds (in this issue) challenged my thesis that the regression-based discrepancy method (RDM) is not a valid tool to detect aptitude–achievement discrepancies. In this response, I show that the statistical and theoretical counterarguments of Willson and Reynolds are based on a misreading of the statistical models presented. Furthermore, I demonstrate that the regression adjustment, which is largest for lower correlations, is the direct source of the lack of validity of the RDM procedure. Nevertheless, RDM can be considered a valid method to measure an achievement component that is unrelated to intelligence.

In my original article (Van den Broeck, in this issue), I argued that the regression-based method to operationalize aptitude–achievement discrepancies is logically inconsistent with the underlying underachievement concept. This argument was framed within an if–then logic: *If* one chooses to endorse the concept of underachievement for one or more domains of academic achievement, *then* the regression-based discrepancy method (RDM) is an invalid diagnostic procedure and is clearly inferior to the simple standard score difference method (SDM). Willson and Reynolds (this issue) have tried to refute my argumentation and conclusions. Essentially, they claim that my initial assumptions about the relationship between intelligence and achievement are inconsistent with contemporary models, whereas the logical derivations based on these assumptions are said to be "incontrovertible" (Willson & Reynolds). To me, this is a surprising statement, because my basic model (see Equations 5 and 6 in my original article) is theoretically almost perfectly neutral and only assumes a positive correlation of whatever size between intelligence and achievement. Thus, this model is easily capable of encompassing any existent theoretical model of the intelligence–achievement constellation, including the model proposed by Willson and Reynolds (see their Figure 1). Moreover, the most specific model I discussed (the model presented in Figure 5 in the original article) is exclusively based on the underachievement concept, which gave rise to the very idea of discrepancy measurements. Willson and Reynolds seem to think that our mutual controversy is in the assumptions, not in the derivations that follow. As the assumptions I took as a starting point for the mathematical derivations about the component structure of the intelligence–achievement relationship *are* the assumptions implied by the underachievement concept, our controversy is not about these assumptions. The real question at issue is the exactitude of the derived conclusions. If they are correct, which Willson and Reynolds do not dispute, they entail important implications for the diagnostic practice of, and the research on, learning disabilities. In this response, I shall try to show in detail that Willson and Reynolds' reading of my original article is based on an unfortunate misunderstanding of the presented material and that their arguments cannot be used to vindicate the alleged superiority of the regression-based discrepancy method.

Learning Disability as Discrepancy: Modeling and Statistical Considerations

In evaluating the measurement models I discussed, Willson and Reynolds mention several models that were previously presented in the literature. For a clear understanding, it will be useful that I shortly explicate, at the outset, the models summarized by Reynolds (1984). Model 1 is the simple difference score between a standardized aptitude score and an achievement score (what I called SDM) without a correction for the unreliability of both measures. Model 2 is McLeod's (1979) version of the regression-based discrepancy method, which takes the difference between an IQ-based predicted achievement score and the observed achievement score as a measure of discrepancy (RDM). The only difference between this model and Model 3 is the peculiar formula used in Model 2 for the determination of the criterion value that has to be exceeded in order to be considered a real discrepancy. As this formula cannot be clearly mathematically determined, this model is currently deemed obsolete. The correct and commonly used criterion value is presented in my Equation 1 and characterizes Model 3. Model 4 is a variant of Model 1 that takes into account the unreliability of the measurements. This model is conceptually and mathematically equivalent to my formulation of SDM, when the reliability of the discrepancy score is determined (see Appendix B in my original article). In short, the critical comparison I made was between Model 3 and Model 4. Therefore, it came as a surprise that Willson and Reynolds interpreted my Equations 5 and 6 as a formulation of Model 4, whereas in fact these equations do not represent any discrepancy measurement model at all; they only express an IQ score and an achievement score as a linear combination of a common factor and a unique factor. In other words, they represent a *generic* model that can be used to evaluate the two discrepancy measurement models I formulated in Equations 1 and 3. Willson and Reynolds apparently identify the C component as the "true" score. This is not what I meant. The C component stands for the common variance of the IQ and achievement measurements. The unique components X and Y include the error variances but are not identical with the error variances. Thus, any individual has a certain C score (C_i) that by definition partly accounts for his or her performance in the IQ and achievement test; this is exactly the meaning of a common score. However, the influence of C on IQ and achievement is not necessarily equal; it depends on the regression weights of C_i. The only case where the C component is eliminated by differencing the IQ and achievement scores is when both regression weights of C are assumed to be equal and a simple difference score (SDM) is used. But even then an SDM score represents at least a difference between an intelligence-specific factor (X) and an achievement-specific factor (Y), whereas in an RDM score the influence of X is, on average, eliminated by a negatively weighted C component (see Figure 4 in my original article). The most interesting case, however, from the point of view of the underachievement concept, is when the regression weights are not equal. In that case, C_i is indeed the true intelligence score (which now can be viewed as true learning potential). Then X represents error variance, and Y stands for error variance plus a specific achievement component. As I have deduced, it is again the SDM score that is superior in detecting an aptitude–achievement discrepancy.

A second point raised by Willson and Reynolds concerns the meaning of regression itself. The authors suggest that their interpretation of regression as "conditional expectation" would be different from my interpretation in terms of "change scores." However, both interpretations are conceptually and mathematically identical, except that change scores terminology is used when the same psychological attribute changes over time within an individual. Thus, there is no quarrel about the meaning of regression. The real point of contention is the relevancy of a regression adjustment when determining an aptitude–achievement discrepancy. According to Reynolds (1984), the reason for the discrepancy emphasis of the legal definition is clear: "The only consensus regarding the characteristics of this 'thing' called learning disability, was that it resulted in a major discrepancy between what you would expect academically of learning disabled children and the level at which they were actually achieving" (p. 452). This is the crucial question; what would we academically (say in reading) expect from a child with an IQ of, for instance, 130? According to the theory of learning potential underlying the concept of underachievement, one would expect a reading score commensurate with the true IQ score (with an IQ reliability of .90, the expected achievement score would be 128.5). However, based on the real correlation between IQ and reading (suppose this is .50), one would expect a reading score of 115. Forced to choose between theory and empirical reality, the choice would be self-evident; this is the delusive appeal of the regression-based discrepancy method.

But even when the theoretical model is forced to take into account the empirical correlation between IQ and achievement (see Figure 5 in my original article), a discrepancy score only makes sense as a measure of the extent that the achievement score departs from the aptitude level, which now is a determinant of minor importance. A regression-based discrepancy score is unfit to detect this discrepancy for the simple reason that it largely adjusts for the influence of the achievement-specific factors that are the very cause of the aptitude–achievement departure. In other words, RDM partly destroys what it aims to measure. Willson and Reynolds ask what we want to know about the difference. The answer is indeed straightforward; we want to know how large the aptitude–achievement difference is. The *description* of this difference is obviously something else than the *explanation* or the *prediction* of the achievement score, in which case a regression equation would be appropriate. In fact, it is RDM that has to be criticized for taking no account of the empirical reality, because it is based on a counterfactual assumption. The regression adjustment term in Equation 2 (see my original article) is directed toward the question, if person i had a mean IQ instead of an IQ of 130, what would have been the observed discrepancy for this person? Formulated for the entire sample, RDM scores seek to determine what the observed discrepancy would have been if everyone had the same IQ score.

This discussion about use and misuse of residual discrepancy scores is closely related to the discussion about the use of residual change scores in the analysis of covariance procedures comparing experimental or nonequivalent groups. It has now been established that the often cited deficiencies of the difference score—low reliability and negative correlation with initial status—are more illusory than real (see Rogosa, Brandt, & Zimowski, 1982). According to Rogosa et al., "the crucial message is that residual change measures are not a replacement or substitute for the estimation of the true change for each individual" (p. 740). In the words of Cronbach and Furby (1970), "one cannot argue that the residualized score is a 'corrected' measure of gain, since in most studies the portion discarded includes some genuine and important change in the person" (p. 74).

Concerning the reliability of the two discrepancy measurement methods discussed, it can be shown that the reliability of SDM is somewhat lower than the reliability of RDM when the standard deviations and reliabilities of the IQ and achievement test are the same (see Zimmerman & Williams, 1982). Assuming test reliabilities of .90, and an IQ–achievement correlation of .50, the reliability of SDM scores is .80, and the reliability of RDM scores is .83 (see Willson and Reynolds, 1984, for the reliability formulas). However, the most important characteristic of a measure is its validity to measure the concept it was designed to measure. Because a discrepancy score aims to measure the discrepancy between learning potential and academic achievement, the respective correlation coefficients of the SDM and RDM scores with the difference between the C component (learning potential) and the true achievement score ($C - AS_t$) have to be determined (see Van den Broeck, 2001b). As shown in Figure 1, SDM is more appropriate than RDM in detecting the discrepancy over the entire range of the IQ–achievement correlation. Only when this correlation is very high—and when consequently the reliability of the discrepancy scores decreases—the correlation coefficients with $C - AS_t$ are almost identical. The figure clearly demonstrates that the divergence between both procedures in detecting discrepancies increases with a diminishing correlation, implying a larger regression correction. Thus, the regression adjustment is directly responsible for the lack of efficiency of the RDM procedure.

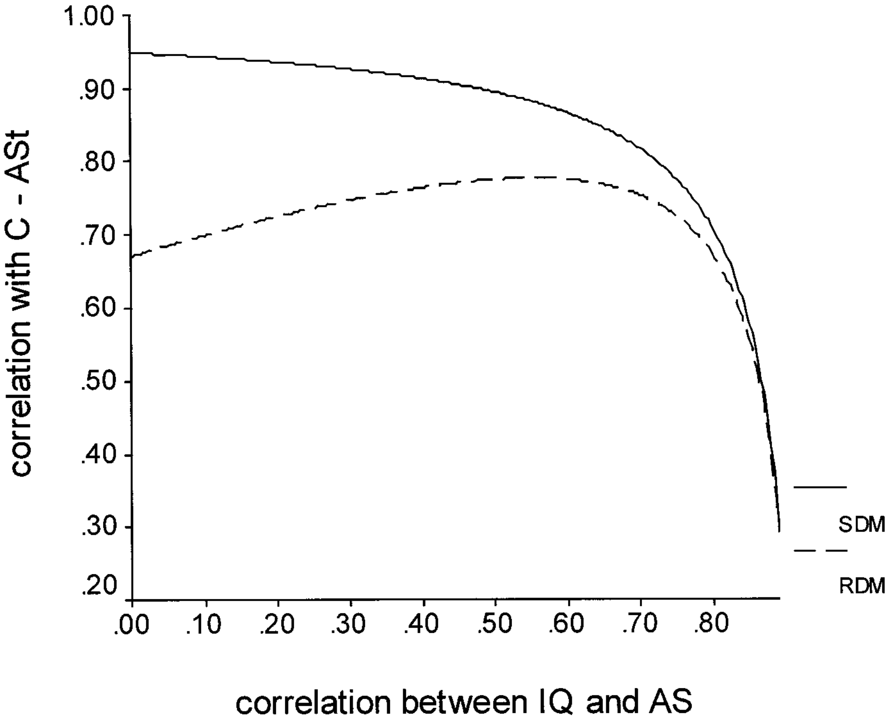

FIGURE 1. Correlation of simple standard discrepancy (SDM) scores and regression-based discrepancy (RDM) scores with the difference between learning potential and true achievement score ($C - AS_t$) as a function of IQ–achievement correlation, assuming an IQ reliability of .90.

According to Willson and Reynolds, our simulation produced exemplar cases that are not realistic in practice. How could that be? Because that simulation study was based on reasonable and realistic assumptions (i.e., normal distribution and IQ–achievement correlation of .50), it produced 8,000 realistic cases. The authors' arguments in terms of legal considerations, realism of IQ levels, and negatively skewed distributions offer, as far as I can see, no real argument in favor of one or the other discrepancy method. The conceptual validity of a measure is not only of theoretical interest, but also of practical importance.

I have argued and mathematically proven that RDM is not a valid measure of aptitude–achievement discrepancy. Nevertheless, RDM is not a worthless measure because it measures something else than hitherto thought. It offers a valid measurement of the unique achievement component (Y)—that is, reading (dis)ability adjusted for the influence of intelligence. The validity of RDM to measure this specific achievement component is generally high (see Van den Broeck, 2001b). Because the correlation between RDM and Y is negative, a positive RDM score indicates a below-average unique achievement score and vice versa. Assuming reliabilities of .90 and an IQ–achievement correlation of .50, this correlation is –.91. For example, for Case 1 (see Table 1), we can infer from the SDM score that this individual is reading about 1.5 standard deviations below his or her intellectual potential, although intelligence is only a minor determinant of reading. Furthermore, the RDM score indicates that the specific (i.e., intelligence-unrelated) reading ability of this individual is only slightly below average. As can be seen, the combination of

TABLE 1
Test Scores, Discrepancy Scores, and Component Scores of 20 Participants From the Simulation Study

Participant	IQ	AS	C	AS_t	Y	SDM	RDM	Y'	$C - AS_t$	DIS_{SDM}	DIS_{RDM}	DIS_{crit}	$Error_{SDM}$	$Error_{RDM}$
1	139	113	146	111	82	25	7	93	37	yes	no	yes	12	30
2	136	106	130	110	92	30	14	86	21	yes	no	no	9	7
3	134	113	136	121	101	21	4	96	16	no	no	no	5	12
4	132	102	131	109	90	30	16	84	23	yes	no	yes	7	7
5	120	93	118	94	81	27	20	80	25	yes	no	yes	2	5
6	119	96	110	101	95	24	16	84	9	yes	no	no	15	7
7	115	85	115	82	68	31	27	73	35	yes	yes	yes	4	9
8	114	88	111	96	87	26	22	78	16	yes	no	no	10	6
9	113	87	114	85	73	26	23	77	30	yes	yes	yes	4	8
10	106	82	107	84	76	24	24	76	25	yes	yes	yes	1	1
11	103	78	100	81	77	25	27	73	20	yes	yes	no	5	7
12	103	81	97	93	93	23	24	76	4	yes	yes	no	18	20
13	100	79	98	75	71	21	25	75	25	no	yes	yes	3	0
14	99	80	96	81	80	19	22	78	16	no	no	no	3	6
15	98	79	96	79	77	19	23	77	18	no	yes	no	1	5
16	96	75	96	78	77	21	27	73	18	no	yes	no	3	8
17	96	75	91	77	78	22	27	73	15	no	yes	no	6	12
18	85	66	84	65	69	19	30	70	20	no	yes	no	1	11
19	76	69	75	66	76	7	22	78	9	no	no	no	2	13
20	69	59	63	63	80	10	29	71	0	no	yes	no	10	29

Note. $N = 8{,}000$. AS = achievement score; C = learning potential or true IQ score; AS = true achievement score; Y = unique component score for achievement; SDM = simple standard discrepancy score; RDM = regression-based discrepancy score; Y' = predicted unique component score for achievement; $C - AS_t$ = difference score between learning potential and true achievement. All test and component scores have a mean of 100 and a standard deviation of 15. All discrepancies (SDM, RDM, and $C - AS_t$) have a standard deviation of 15. The predicted Y score (Y') is directly derived from the RDM score and has a correlation with it of –1. All categorical decisions of discrepancies (DIS) are based on a critical value of 1.5 *SD*. DIS_{crit} is based on $C - AS_t$. All numbers are rounded to the nearest integer.

both measures yields some interesting information and a realistic appraisal of the role of intelligence-related and intelligence-unrelated determinants of achievement.

IQ–Achievement models

Finally, Willson and Reynolds' critique on my so-called conceptual model of intelligence and achievement is entirely based on their misreading of the expressions of IQ and achievement as linear combinations of a common component and unique components (see my Equations 5 and 6). None of the models I presented assume that achievement is exclusively determined by the common factor, as these authors seem to think. On the contrary, in my second model (see my original Figure 5), achievement is primarily determined by an achievement-specific factor (Y), including the error variance. As already argued, Willson and Reynolds' model of the IQ–reading constellation could be easily encompassed within this second model, except that their model is statistically misspecified because reading is exclusively determined by error variance and, hence, remains unexplained. Their statement that the manifest variables in our model are correlated independently of the latent direct effects may be the result of my unusual addition of curved arrows in my original Figure 3 and Figure 5 between dependent variables. Despite common path-analytical usage to interconnect only correlated independent variables, I added a curved arrow between IQ and achievement for a better understanding of the less initiated reader in path analyses. The IQ–achievement correlation is, as explained in Appendix A of my original article, the direct result of the influence of the latent variables.

Actually, I sympathize with Willson and Reynolds' description of the reading process as determined by more or less specific cognitive factors (see also Van den Broeck, 2001a). The point is that the domain specificity of word reading empirically falsifies the assumption of a dominant role of intelligence underlying the underachievement concept. As a consequence, the crucial role of an intelligence–reading discrepancy in the definition of reading disability or dyslexia cannot be justified. This doesn't preclude, however, a more modest role of intelligence–reading discrepancies in the assessment of reading problems. As exemplified here, it is possible to estimate validly the respective influences of intelligence-related and reading-specific determinants of a reading score by making use of discrepancy measures. In conclusion, my position that RDM offers an invalid and biased estimate of aptitude–achievement discrepancies remains unrefuted by the critiques of Willson and Reynolds.

ABOUT THE AUTHOR

Wim Van den Broeck, *PhD, is a researcher at the University of Leiden. He is interested in the* *cognitive processes involving normal and disabled reading. He was trained as an experimental psychologist and is currently lecturing in methodology. Address: Wim Van den Broeck, Department of Social Sciences: Section of Special Education, University of Leiden, Wassenaarseweg 52, 2300 RB Leiden, The Netherlands; e-mail: broeck@fsw.leidenuniv.nl*

REFERENCES

Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—or should we? *Psychological Bulletin, 74,* 68–80.

McLeod, J. (1979). Educational underachievement: Toward a defensible psychometric definition. *Journal of Learning Disabilities, 12,* 42–50.

Reynolds, C. R. (1984). Critical measurement issues in learning disabilities. *The Journal of Special Education, 18,* 451–476.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92,* 726–748.

Van den Broeck, W. (2001a). *The concept of developmental dyslexia: Toward a behaviorally grounded diagnosis.* Manuscript submitted for publication.

Van den Broeck, W. (2001b). *The reliability of aptitude–achievement discrepancy measures.* Manuscript in preparation.

Willson, V. L., & Reynolds, C. R. (1984). Another look at evaluating aptitude–achievement discrepancies in the diagnosis of learning disabilities. *The Journal of Special Education, 18,* 477–487.

Zimmerman, D. W., & Williams, R. H. (1982). The relative error magnitude in three measures of change. *Psychometrika, 47,* 141–147.