Learning and Individual Differences 18 (2008) 308–315

Contents lists available at ScienceDirect

Learning and Individual Differences

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

The importance of measuring growth in response to intervention models: Testing a core assumption ☆

Christopher Schatschneider [a,c,*], Richard K. Wagner [a,c,*], Elizabeth C. Crawford [b,c,*]

[a] Department of Psychology, Florida State University, 1107 W. Call Street, Tallahassee, Florida 32306-4301, USA
[b] Department of Communication Disorders, Florida State University, Tallahassee, Florida 32306, USA
[c] Florida Center for Reading Research, Tallahassee, Florida 32301, USA

ARTICLE INFO

Article history:
Received 7 January 2008
Received in revised form 10 April 2008
Accepted 14 April 2008

Keywords:
RTI
Identification
Reading
Dyslexia

ABSTRACT

A core assumption of response to instruction or intervention (RTI) models is the importance of measuring growth in achievement over time in response to effective instruction or intervention. Many RTI models actively monitor growth for identifying individuals who need different levels of intervention. A large-scale (N=23,438), two-year longitudinal study of first grade children was carried out to compare the predictive validity of measures of achievement status, growth in achievement, and their combination for predicting future reading achievement. The results indicate that under typical conditions, measures of growth do not make a contribution to prediction that is independent of measures of achievement status. These results question the validity of a core assumption of RTI models.

> "The greatest enemy of the truth is not the lie-deliberate, contrived, and dishonest, but the myth-persistent, pervasive, and unrealistic." John F. Kennedy

1. Introduction

Reading disability or developmental dyslexia refers to unexplained poor performance in reading. When the concept of reading disabilities was formalized in the 1960's, the common assumption made when children were unable to learn to read was that they were impaired intellectually. Later, it was recognized that an inability to learn to read could exist despite the absence of general intellectual impairment (Kirk, 1962). As a result, the traditional operational definition of reading disability in the United States was rooted in a comparison of performance in reading and performance on a measure of cognitive ability. In addition to requiring that the observed poor performance in reading could not be explained as a result of a general intellectual impairment, other potential explanations that needed to be ruled out included lack of an opportunity to learn and impaired sensory capacities in areas required for reading, especially vision (Lyon, Shaywitz, & Shaywitz, 2003).

The need to distinguish among low-achievers in reading is rooted in the acknowledgement that there are multiple causes of poor reading skills, including low general cognitive ability, minimal opportunities to learn how to read at home (or perhaps also at school), sensory impairments, or possibly the existence of some specific neurological impairment. Presumably the identification of specific causes would then lead to specific treatments for poor reading. The traditional approach in looking for children with specific neurological impairments in reading has been to use adequate overall intellectual function as a proxy to rule out these other causes of low reading achievement.

However, the traditional approach to identification has been challenged over the years on a number of grounds. The most recent challenge has come from proponents of an approach to identification of individuals with reading disability on the basis of their failure to respond to provision of effective instruction and intervention. In the present article, we (a) review concerns that have been raised about the traditional approach to identification, (b) review approaches to identification based on response to instruction and intervention, (c) present results from a large-scale, longitudinal study comparing the predictive validity of measures of achievement status, growth in achievement, and their combination for predicting future reading achievement, and (d) consider implications of the results for evaluating the potential of response to intervention models for addressing limitations associated with traditional models.

☆ Support for this research has been provided by Grant P50 HD052120 from NICHD.
* Corresponding authors.
E-mail addresses: schatschneider@psy.fsu.edu (C. Schatschneider), rkwagner@psy.fsu.edu (R.K. Wagner), ecrawford@fcrr.org (E.C. Crawford).

1041-6080/$ – see front matter © 2008 Published by Elsevier Inc.
doi:10.1016/j.lindif.2008.04.005

2. The discrepancy approach to identification

The most common approach for determining whether a child is eligible for special education services in the United States due to the presence of a specific learning disability in reading has required evidence of a discrepancy between IQ and reading performance. This IQ-achievement discrepancy approach has come under attack for three major reasons (Wagner, in press). These include viewing the approach as a "wait to fail" model, having concerns about the reliability of IQ-achievement discrepancy scores, and questioning the validity and educational relevance of the distinction between poor readers who are IQ-discrepant and poor readers who are not (Fletcher, Lyon, Fuchs, & Barnes, 2007; Fletcher, Francis, Rourke, Shaywitz, & Shaywitz, 1992; Fletcher, Morris, & Lyon, 2003; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Lyon et al., 2003; President's Commission on Excellence in Special Education, 2002; Siegel, 1992; Spear-Swerling & Sternberg, 1996; Stanovich, 1991; Stanovich & Siegel, 1994; Stuebing et al., 2002.). We briefly review each of these concerns about the IQ-achievement discrepancy model.

2.1. A "wait to fail" model?

Under the traditional approach to identification, it indeed has been the case that most children with reading disability have not been identified and provided additional services until second grade. This is problematic to the extent that reading problems become more intractable the longer they exist. There is support for the idea that reading problems become increasingly intractable from the results of longitudinal correlational studies of the development of reading. Individual differences in reading skills become remarkably stable by second grade (Francis et al., 1996; Wagner et al., 1997). For example, Wagner et al. (1997) reported year-to-year correlations for word-level decoding of .84, .96, and .96, for the time periods 1st to 2nd grade, 2nd to 3rd grade, and 3rd to 4th grade respectively, in their five-year longitudinal study of over 200 children. A related concern of a "wait to fail" approach is that a child who struggles with early reading may experience unfortunate concomitants such as a negative academic self-concept or an aversion to schooling.

There are at least three reasons for the fact that most children with reading disability are not identified and provided additional services until second grade. First, reading instruction in the United States begins in earnest in first grade when children typically are six years old. This marks the beginning of compulsory education. Although there is a trend of teaching more about reading in the previous kindergarten year and even in preschools, there is great variability in the educational experience of children prior to first grade. Kindergarten typically is optional, and may exist as either a half- or full-day program. Preschools vary tremendously in the extent to which early literacy activities are emphasized. If reading instruction does not begin in earnest until first grade, any approach that is based on an observed failure to acquire reading skills that are taught formally in school cannot hope to identify children until some time after such instruction has begun. The second, related reason is that common measures of reading achievement show floor effects through the beginning of first grade, as might be expected if they assess reading skills that are taught in first grade. These floor effects make it virtually impossible to observe a discrepancy between aptitude and achievement much before the end of first grade or early second grade. The third reason is that resources limit the number of students who can be referred for evaluation, which adds an additional time lag into the provision of services.

There is no disputing the fact that the traditional IQ-achievement discrepancy approach typically does not result in getting help to children with reading disability until second grade on average, and this is problematic. However, new developments in our understanding of early literacy potentially could eliminate or reduce the "wait-to-fail" nature of the traditional approach. Reading should not be thought of as a skill that appears out of whole cloth when children are first taught to read. Rather, reading is better conceptualized as a developmental phenomenon that builds on early print awareness, rudimentary phonological awareness, and vocabulary. These skills can be assessed reliably in children as young as 3 years of age, and are highly predictive of later decoding (Burgess & Lonigan, 1998; Lonigan, Wagner, Torgesen, & Rashotte, 2007). If measures of print awareness replace measures of reading achievement, it should be possible to identify children whose print awareness is discrepant from their IQ prior to their entry in kindergarten or first grade. Of course, it would be important to evaluate whether early IQ-print awareness achievement discrepancies are stable and become IQ-reading achievement discrepancies at a later time.

2.2. Are IQ-achievement discrepancy scores unreliable?

In practice, an operational definition of an IQ-achievement discrepancy is required if it is to be useful for identification of individuals with reading disability. For example, an operational definition might be a difference of 15 standard score points between accepted measures of cognitive ability and of reading. Under some circumstances, differences scores can be unreliable. The formula for calculating the reliability of a difference score is the following:

$$r_{\text{diff}} \frac{[(r_{aa} + r_{bb})/2] - r_{ab}}{1 - r_{ab}}$$

In this formula, r_{diff} is the reliability of the difference score, r_{aa} and r_{bb} are the reliabilities of the two scores used to create the difference score, and r_{ab} is the correlation between these two scores. When the two scores that comprise the difference score are uncorrelated, the reliability of the difference score equals the average of the reliabilities of the two scores. This can be seen in the formula above by using zero as the value of r_{ab}. The numerator simplifies to $(r_{aa}+r_{bb})/2$ (i.e., the average of the reliabilities of the two scores) and the denominator simplifies to one. But now consider what happens as the correlation between the two scores begins to approach their reliabilities. The numerator approaches zero as r_{ab} approaches $(r_{aa}+r_{bb})/2$. In the extreme case, when the correlation between the two scores is equal to their average reliability, the numerator becomes zero, and consequently, the reliability of the difference score becomes zero. For most common measures of IQ and achievement, their reliabilities are quite high but they are correlated substantially. Therefore, the reliability of IQ-achievement differences tends to be less than the reliability of typical IQ or achievement scores but certainly greater than zero.

Fletcher et al. (2007) identified two additional problems that reduce the reliability of IQ-achievement difference scores. First, when an IQ-achievement discrepancy score is used for identification, a cut-off score must be identified. Because the distributions of IQ and achievement scores are continuous, so is the distribution of IQ-achievement discrepancy scores. Consequently, an arbitrary cut-point is imposed on a continuous distribution and this can have a detrimental effect on reliability. Reliability is reduced because individuals who score close to the cut-point are likely to vary on which side of the cut-point they land upon repeated testing. In an empirical investigation of this phenomenon using both real data from the Connecticut Longitudinal Study and simulated data, Francis et al. (2005) reported that classification decisions showed instability longitudinally that could be explained in part to the imposition of arbitrary cut-points on a continuous distribution.

The second problem identified by Fletcher et al. (2007) that could affect the reliability of classification decisions based on IQ-achievement discrepancy scores is regression to the mean. Regression to the mean refers to the fact that when individuals are selected on the basis of an extreme score on a test, repeated testing with any test that is