Published in final edited form as:

Annee Psychol. 2014 December 1; 114(4): 753–777. doi:10.4074/S0003503314004072.

Issues in Identifying Poor Comprehenders

Janice M. Keenan,
University of Denver

Anh N. Hua,
University of Denver

Chelsea E. Meenan,
University of Denver

Bruce F. Pennington,
University of Denver

Erik Willcutt, and
University of Colorado, Boulder

Richard K. Olson
University of Colorado, Boulder

Abstract

Studies of poor comprehenders vary in the selection criteria and tests that they use to define poor comprehension. Could these differences play a role in determining findings about poor comprehension? This study assessed the extent to which differences in selection methods affect who gets identified as poor comprehenders, and examined how their cognitive profiles differ. Over 1,500 children, ages 8 – 19, took multiple tests of reading comprehension, listening comprehension, single word reading and nonword reading. Poor comprehension was defined by performing in the low-tail and by discrepancies either with word or nonword reading. Odds of any two selection methods identifying the same individuals were generally low, and depended on type of comprehension test more than modality, as well as selection criteria, and comprehender's age. Poor comprehenders selected by the different methods were found to vary in IQ, working memory, but not attention. The findings show that differences across studies in tests and selection criteria used to define poor comprehension are not insignificant and can have substantial consequences for what is meant by poor comprehension and its associated deficits.

1. Introduction

Research on reading comprehension has increasingly focused on identifying individuals who consistently struggle with comprehension, referred to as *poor comprehenders* (Hulme & Snowling, 2011). By studying how poor comprehenders differ from typical readers, researchers have sought to better understand the nature of comprehension processes, the

Correspondence concerning this article should be sent to Janice M. Keenan, Department of Psychology, University of Denver, 2155 S. Race, Denver, Colorado 80208. jkeen@du.edu.

Portions of these data were presented at the Society for the Scientific Study of Reading in Montreal in July, 2012.

causes of comprehension failures, and the stability of comprehension disorders (e.g., Cain & Oakhill, 2006, 2007; Nation & Snowling, 1998; Elwér, Keenan, Olson, Byrne, & Samuelsson, 2013; Oakhill, Yuill, & Parkin, 1986). Because they often go unrecognized in the classroom, identifying poor comprehenders has become an important goal for educators as well since academic success depends on having good reading and oral comprehension. If identification of poor comprehension can be made early, intervention is likely to be more effective (Bianco, Pellenq, Lambert, Bressoux, & Doyen, 2012; Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013).

A potentially complicating factor for both researchers and educators, however, is that the methods used for identifying poor comprehenders vary considerably. Some of this variability in selection criteria is motivated by theoretical considerations. This is the case when some researchers choose to select those who are not just poor on comprehension, but poor on comprehension despite adequate having decoding skill (*specifically poor comprehenders*), whereas others may study all poor comprehenders. Some variability is also dictated by one's goal; whereas clinicians may focus on all poor comprehenders, some researchers may choose to be more *specific* in their focus. Other differences in selection arise for practical reasons, such as the type of comprehension test one uses; tests vary across nations and languages, but even within nations, there are usually many options, with differences occurring in factors such as the length of passages, subject matter of passages, type of comprehension assessment (e.g., cloze test, multiple-choice questions, open-ended questions, retelling).

The result is that there can be considerable differences across studies of poor comprehenders both in the specifics of how comprehension is assessed and in the selection criteria used for defining poor performance. The specific tests used for comprehension assessment can vary in the modality in which comprehension is being assessed (reading or oral), as well as in attributes such as those already noted that can affect the component skills assessed. The selection criteria used for defining who is a poor comprehender can be simply those performing below a cutoff (which itself varies across studies), or poor comprehension can be defined relative to the child's decoding, which is variously assessed by either word or nonword reading. Furthermore, because age interacts with decoding skill, the same selection criteria may result in greater differences between modality and between tests depending on the comprehender's age.

The current research therefore asks: How much do these differences in method of identification matter in the diagnosis of poor comprehension? By directly comparing different methods of selection of poor comprehenders on the same sample of children, we assess the consequences of particular selection criteria for identifying who is a poor comprehender, both for younger and for older children. Specifically, we ask to what extent do the same children get identified across different methods. This is an important question because the research literature grows by compiling findings across studies. If we find consistency in identification across methods, it will provide justification for this cross-study aggregation. If we find inconsistencies, it will still be quite informative, because knowing that selection factors can alter the types of comprehenders in the samples can help explain why conclusions about poor comprehenders might vary across studies.

2. Prevalence of Poor Comprehenders

In their review of research on children's reading comprehension difficulties, Hulme and Snowling (2011) discuss the prevalence of reading comprehension difficulties and conclude that "there is little doubt they are relatively common". Specific values for prevalence rates in primary-school children who struggle to comprehend texts that they can accurately decode are often said to be around 10% (e.g., Clarke, Snowling, Truelove, & Hulme, 2010; Nation & Snowling, 1998).

Prevalence statistics suggest that there is a clearly definable category of comprehension deficit. It is important to note, however, that no matter what assessment instrument is used, comprehension skills show continuous variations within the population. There is no break in the distribution separating poor comprehenders from others. The continuous nature of the distribution of comprehension skill has two important implications for assessments of prevalence of comprehension disorders. One is that we recognize that any diagnostic cut-off values used in the identification of poor comprehenders are essentially arbitrary, and consequently, can vary across studies. Second, because estimates of the prevalence rate for poor comprehension depend on the cut-off value used, prevalence rates are equally arbitrary. Thus, our investigation of the degree of consistency in identification across selection methods will demonstrate how much prevalence statistics depend on the selection criteria used to define poor comprehension.

3. Selection Criteria for Defining Poor Comprehenders

3.1. Reading Comprehension Tests

Selecting from among the wide variety of choices for a comprehension test to use in research or diagnosis is often based on factors such as how long the test takes to administer and how easy it is to score. The availability and popularity of the test also figures in people's selection. A rather popular test for assessing reading comprehension in studies of poor comprehenders in Britain is the Neale Analysis of Reading Ability (NARA). One popular feature of this normed test is that it conveniently assesses word decoding at the same time as assessing comprehension on each passage, thus saving the child from having to take additional decoding tests and saving the researcher valuable time. Although the decoding and comprehension measures are frequently interpreted as separate assessments, the separability is not clear-cut. It has been shown that the NARA underestimates comprehension in children with weak decoding skills (Spooner, Baddeley, & Gathercole, 2004). Thus, because of the dependence of comprehension on decoding skills, it raises the question of how decoding skill plays into the identification of poor comprehenders.

The question of how decoding skill affects the identification of poor comprehenders is not just an issue for tests that do combined assessment like the NARA. It has become an issue for all comprehension assessment because we now know that test format differences that affect ease of administering and scoring also create differences in the skills being assessed by the test (Cutting & Scarborough, 2006; Francis, Fletcher, Catts, & Tomblin, 2005; Keenan, Betjemann, & Olson, 2008; Nation & Snowling, 1997). By comparing the same readers on different reading comprehension tests, these studies have found that tests differ in

the extent to which their variance is explained by decoding skills versus oral language skills. As a result, a reader can appear to have poor comprehension when assessed with one test, but not when assessed with another test (Keenan & Meenan, 2014). Not only do reading comprehension tests differ in who they classify as being in the low-tail, they also differ in who they identify as top performers (Keenan & Meenan, 2014).

Some reading comprehension tests have almost all of their variance explained by word decoding skill (Cutting & Scarborough, 2006; Francis, Fletcher, Catts, & Tomblin, 2005; Keenan, et al., 2008; Nation & Snowling, 1997). Keenan et al. (2008) proposed that these are tests that use single sentence-texts or very short texts with little contextual support for identifying words; comprehension hinges mostly on correct decoding in these tests because the passages are too short to involve the skills required to construct mental models of situations that dynamically change across a passage.

The tests that load more highly on a decoding factor than on a comprehension factor have been labeled *RC-D* tests, whereas those that load more highly on comprehension than on decoding are called *RC-C* tests (Betjemann, Keenan, Olson, & DeFries, 2011). The difference between *RC-D* and *RC-C* tests is particularly important from a developmental perspective. Decoding accounts for more variance in reading comprehension when children are younger, and their word decoding skills are weaker, than it does in older children, for whom listening comprehension explains more of the variance (Catts, Hogan, & Adlof, 2005; Curtis, 1980; Vellutino, Tunmer, Jaccard, & Chen, 2007). However, this developmental shift seems to be greater for *RC-D* than *RC-C* tests (Christopher, et al., 2012; Keenan, et al., 2008). Consequently, in our analyses of selection methods we will compare these two types of assessments across the full age range, as well as separately for younger and for older comprehenders.

3.2. Listening Comprehension Tests

Because separating poor comprehension from the effects of poor decoding is difficult, particularly for younger readers, some researchers have opted to assess comprehension with oral language instead of reading when selecting poor comprehenders (e.g., Elwer, et al., 2013). Others have used listening comprehension as an initial selector and then have further selected poor readers from that sample (e.g., Clarke, et al., 2010). Still others have done the reverse, using listening comprehension secondarily to select from among those identified as poor readers (e.g., Catts, Hogan, & Fey, 2003).

It is an open question as to how similar poor comprehenders selected on the basis of a listening test might be to those selected by a reading test. Listening comprehension provides an assessment of comprehension skill uncontaminated by decoding, so to the extent that a reading comprehension test is an *RC-D* test, there may be considerable differences in who is identified as a poor comprehender when using listening versus reading. It should be noted, however, that if poor comprehenders are selected by poor listening comprehension, it does not mean that these children do not also have poor decoding; word decoding and comprehension are highly correlated ($r = .73$ in our sample). So if one wants to select poor comprehenders on the basis of listening comprehension performance who do not have poor decoding skill, it is necessary to also use a discrepancy criterion for selection.

There may be other differences in the cognitive requirements of listening comprehension and reading comprehension than just decoding. For example, attention difficulties may impact listening more than reading because of the transient nature of the input. In reading, lapses of attention can be remedied by rereading, whereas in listening comprehension, there is no way to recover unattended input. For all these reasons, then, it is of interest to know whether poor comprehenders defined by listening comprehension differ from those defined by a reading comprehension test.

3.3. Cut-Offs for Defining Poor Performance

It is common to identify children as having clinically significant comprehension difficulties by specifying a cut-off on a comprehension test to define poor performance. Across studies, poor comprehenders have been defined by various cut-off values. The specific value selected is often dictated by the size of the available sample. If the sample of comprehenders from which one is selecting poor performers is rather small, researchers tend to use a higher percentile in order to increase the number of poor comprehenders to study.

Although everyone uses cut-offs, it is important to recognize that there are some interpretation issues associated with their use. A cut-off value artificially imposes a dichotomous categorization on data that is essentially continuous. It defines those on one side of the cut-off as having a deficit and those on the other side as not; but because of measurement error, the true score might fall on the other side of the cut-off than the observed score. Increasingly, researchers have sought to circumvent the effect of these measurement issues on their conclusions by doing things like examining their results using multiple cut-offs to determine the generalizability of their findings (Elwer, et al., 2013; Keenan & Meenan, 2014) or by assessing the stability of deficits over time (Catts, et al., 2003; 2006; Elwer, et al., 2013; Nation, Cocksey, Taylor, & Bishop, 2010; Oakhill & Cain, 2012), or even doing all analyses continuously rather than categorically (e.g., Keenan, et al., 2006).

3.4. Discrepancy Measures for Defining Specifically Poor Comprehenders

Although difficulties in reading comprehension due to deficits in word decoding have long been recognized (Liberman, Shankweiler, Fisher, & Carter, 1974; Perfetti, 1985; Shankweiler, 1989), what has captured the attention of many researchers in recent years are those cases of comprehension difficulties not associated with decoding problems. These children are referred to as *specifically poor comprehenders* (Oakhill, Cain & Bryant, 2003; Catts, et al., 2003; Oakhill, 1994; Stothard & Hulme, 1992; Yuill & Oakhill, 1991). Partial independence of individual differences in decoding and comprehension has even been established now at a genetic level (Harlaar, et al., 2010; Keenan, Betjemann, Wadsworth, DeFries, & Olson, 2006).

Specifically poor comprehenders are defined as having a discrepancy, typically on the order of one standard deviation, between their decoding skill and comprehension. Because using only a discrepancy in skill can end up selecting as poor comprehenders children whose comprehension is perfectly adequate but whose decoding skill is exceptional, most studies use both a cut-off and a discrepancy (note that the measurement error issues previously

discussed under cut-offs also apply to discrepancies). The typical identification procedure is to use a cutoff to select those in the low-tail of the distribution on comprehension, and then from among those poor comprehenders, select the specifically poor comprehenders whose decoding skill is at least one standard deviation above their comprehension.

The test of decoding skill used for defining the discrepancy in selecting specifically poor comprehenders is another variable that differs across studies. It can be either a test of word reading or a test of nonword reading, or both (e.g., Pimperton & Nation, 2010). Although word and nonword reading are highly correlated, in a language like English with many words having statistically infrequent, or irregular, orthographic-phonological mappings, there is a difference in the role of vocabulary knowledge on word versus nonword reading. Because vocabulary also influences comprehension as well as word reading (Ouellette, 2006), selecting poor comprehenders on the basis of a discrepancy with word reading, as opposed to nonword reading, could make a difference in the extent to which weak vocabulary defines poor comprehension.

4. Current Study

The present article explores the implications of definitional differences for identifying poor comprehenders. Will a child identified as a poor comprehender by one method be similarly identified if different tests or different selection criteria are used?

One might think that a child who performs so poorly as to be in the very low end of the distribution will likely perform poorly no matter how assessed. After all, component comprehension skills are interrelated (Cain & Oakhill, 2007), so a severe deficit in one component could affect all related skills such that the child ends up performing poorly on all tests. However, that is not the case; different reading comprehension tests identify different children as poor (Keenan & Meenan, 2014). Because those differences in identification largely reflect differences in how much variance on the test is explained by word decoding (*RC-D* or *RC-C* tests), it raises the possibility that there may be more consistency in identification between modalities than within. Because *RC-C* tests have so little of their variance explained by decoding skill, they may be more consistent with listening tests, where decoding is irrelevant, than with *RC-D* reading tests, whose variance is largely explained by decoding. We will therefore examine consistency of identifying poor comprehenders both across different reading tests and across modalities. We will examine consistencies both for poor comprehenders defined by performance in the low-tail and specifically poor comprehenders defined by low-tail plus discrepancy with decoding. When examining specifically poor comprehenders, we will see if defining the discrepancy by using word versus nonword reading affects consistency of identification.

We are able to address these questions because, as part of an ongoing behavior genetic study of comprehension (Keenan, et al., 2006; Keenan, Olson, & Betjemann, 2009), we have tested more than 1500 children on a range of both reading and listening comprehension tests. Furthermore, because our study is part of the Colorado Learning Disabilities Research Center, these children are given not only our comprehension tests but also multiple tests of single word reading and nonword reading and a broad assessment of other cognitive skills

(DeFries, et al., 1997; Olson, 2006). We can thus examine the consequences of different definitions of poor comprehension in terms of the consistency with which they identify the same children. Further, we can compare cognitive profiles of the children identified as poor comprehenders across methods to understand what deficits may be varying as a function of selection. Finally, not only can we determine whether there are differences between selection methods in who is defined as poor, but we can also evaluate the impact of age on consistency; because the ages of children in our sample represent a broad range, we can determine whether there is more consistency or less consistency across methods of identification for younger versus older children.

4.1. Method

4.1.1. Participants—The sample consisted of 1,522 children (751 males, 771 females): 1,342 twins¹ and 180 of their siblings recruited for a behavioral genetic study of comprehension skills (Keenan, et al., 2006; 2009) as part of the Colorado Learning Disabilities Research Center (Olson, 2006). The median age was 11.17 years (range 8–19). All were native English speakers. Ethnicity of the sample was: 89% Caucasian, 4.3% Hispanic, 2.2% American Indian, 1.4% Asian, 1.3% African American, and 3% not reported or reported as other. The representativeness of our sample can be further gleaned from Table 1, which presents the standard scores for all those measures we used in the study that are nationally normed, and shows overall performance slightly above average.

4.1.2. Reading Comprehension Tests

RC-D Tests: Based on previous research, two of our reading comprehension measures were known to load more highly on decoding than on comprehension, and thus are called RC-D tests (Betjemann, et al., 2011; Keenan, et al., 2008). One is the Peabody Individual Achievement Test (PIAT, Dunn & Markwardt, 1970) wherein single-sentence texts are read and comprehension is tested by multiple-choice selection of the picture that best expresses the meaning of the sentence. The other is the Woodcock Johnson Passage Comprehension-3 (WJPC, Woodcock, McGrew, & Mather, 2001), which involves reading single-sentence or two-sentence passages and providing a missing word (cloze format) to demonstrate comprehension.

RC-C tests: Previous research showed that our other two reading comprehension tests load more highly on comprehension than on decoding and thus are called RC-C tests (Betjemann, et al., 2011; Keenan, et al., 2008). One is the Gray Oral Reading Test-3 (Wiederholt & Bryant, 1992) involving expository and narrative passages that range in length from 85-150 words and where comprehension is tested with multiple-choice questions. The other is the Qualitative Reading Inventory-3 (Leslie & Caldwell, 2001), which involves reading even longer narrative and expository passages (250-785 words), and where comprehension is assessed both by retelling the passage and answering open-ended comprehension questions. A composite formed from the standardized scores of the retellings and questions was used.

¹It should be noted that if dependence within twin pairs has any effect, it would bias the results in favor of consistency in diagnosis, not inconsistency.

4.1.3. Listening Comprehension Tests—Three listening comprehension tests were used. One was the Woodcock-Johnson Oral Comprehension subtest (Woodcock et al., 2001) in which participants listen to very short texts and provide a missing word to demonstrate their comprehension. Another is a listening version of the QRI-3 (Leslie & Caldwell, 2001) described above under reading which uses multi-paragraph passages. The third is the KNOW-IT Test (Barnes, Dennis, & Kaefele-Kalvaitis, 1996) which has participants listen to a long (7-minute) story and answer open-ended comprehension questions.

4.1.4. Decoding Measures

Word decoding: Word decoding ability was a composite based on *z*-scores from the Timed Oral Reading of Single Words (Olson, Forsberg, Wise, & Rack, 1994) and the PIAT word recognition subtest (Dunn & Markwardt, 1970).

Nonword Decoding: Nonword decoding skill was a composite of two tests of nonword reading developed by Olson et al. (1994). One assessed reading 45 one-syllable nonwords (e.g., *ter*, *strale*). The other assessed reading of 40 two-syllable nonwords (e.g., *vogger*, *strempick*).

4.1.5. Cognitive Profile Measures

IQ: IQ was measured by the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) or Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981).

ADHD symptom count: The Attention Deficit Hyperactivity Disorder Rating Scale-IV (ADHDRS-IV, DuPaul, Power, McGoe, Ikeda, & Anastopoulos, 1998) was used to assess number of symptoms associated with ADHD. Ratings were obtained from both parents and teachers on 18 symptoms. Parent and teacher ratings were combined by positively coding each symptom if it was endorsed by either the parent or the teacher (Lahey et al., 1994; Willcutt, Pennington, Olson, Chhabildas, & Hulslander, 2005).

Working memory: Working memory was measured using a composite of sentence span (Daneman & Carpenter, 1980), counting span (Case, Kurland, & Goldberg, 1982), and digit span using forward and backward digit span from the WISC-R or the WAIS-R.

4.1.6. Procedure

Testing: Tests were administered over two day-long sessions. IQ, word and nonword reading were among the tests administered on the first day. The second day of testing typically occurred within a month of the first session and included all the tests of comprehension, as well as the sentence span and counting span tests of working memory.

Composites and standardization: Composites of all the reading comprehension tests (Read Composite) and of all the listening comprehension tests (Listen Composite) were constructed. In addition to these overall composites, composites were constructed for each specific type of reading comprehension test (RC-D or RC-C Tests). First we standardized scores of each test on our sample of 1,522 children. For all tests, *z*-scores were computed using the standardized residuals that were saved after regressing on children's age and age

squared. We then averaged the standardized scores for the particular tests that were relevant for the specific composite (e.g., the four reading comprehension tests for the Read Composite) and then restandardized those composite scores.

Definitions of Poor Comprehenders: When assessing consistency of identification defined solely by performing in the low-tail of the distribution on a comprehension test, we used the lowest 15% of the sample, which for our sample consisted of 229 children. When assessing consistency of identification of *specifically* poor comprehenders whose comprehension is unexpectedly lower than their decoding skill, we used both a criterion of performing in the low-tail and having a discrepancy. Because decoding and comprehension are highly correlated and thus the cases where they dissociate are infrequent, the cut-off used for the low-tail for the specifically poor comprehenders was higher than that used for low-tail only comparisons; we used the lowest 25%, a fairly common cut-off in studies of specifically poor comprehenders. The discrepancy between decoding and comprehension required for specifically poor comprehenders was that the comprehension *z*-score had to be at least one standard deviation below the decoding *z*-score. We compared prevalence rates as well as consistencies in identification across different selection tests for specifically poor comprehenders both when they were defined by a discrepancy with their word decoding and when they were defined by a discrepancy with their nonword reading.

Assessments of consistency: We compare consistency of identification across modality when poor comprehension is defined solely by performing in the low-tail of the distribution. We compare reading to listening both by examining the consistency when identification is based on a composite over four reading tests, as well as when reading is broken down by type of test, *RC-D* versus *RC-C* tests. We examine these low-tail consistencies both for the full sample, and broken out for younger and older comprehenders. We then examine consistencies across these same methods and age groups for specifically poor comprehenders, both when comprehension skill is defined as discrepant with word reading and when discrepant with nonword reading. Cognitive profiles of the different groups selected by the different methods are then examined by comparing attention, working memory, and IQ scores.

4.2. Results

4.2.1. Correlations—Table 2 presents the correlations, computed across the full sample, among the variables that are used for identifying poor comprehenders. Interestingly, the two components of reading comprehension in the simple view of reading (Hoover & Gough, 1990) correlate equally with the reading comprehension composite ($r = .74$ for listening comprehension, $r = .73$ for word reading). However, the *RC-C* and *RC-D* tests show an imbalance in their correlations with the components. *RC-C* tests show a higher correlation with listening comprehension (.71) than with word reading (.54), whereas *RC-D* tests show the reverse (.64 for listening, .78 for word reading); this provides further support for the distinction between the tests and a reason to consider them separately when evaluating how different types of tests compare in defining poor comprehenders.

Word reading and nonword reading are highly correlated (.83). Both word and nonword reading are more highly correlated with reading comprehension than with listening comprehension, as would be expected. However, word reading correlates more highly than nonword reading does both with reading and with listening comprehension. This likely reflects the shared influences of vocabulary on word identification and on comprehension.

4.2.2. Consistency of Low-tail Identification—Because it is common, especially in clinical and educational settings, to assign a diagnosis of poor comprehender based solely on poor performance on a comprehension test, we first evaluated how consistent our tests were in identifying the poorest performers, defined as the lowest 15% of scorers. Table 3 shows consistency of identification for the reading composite compared to the listening composite, for each type (*RC-D*, *RC-C*) of reading comprehension test compared to the listening composite, and across *RC-D* and *RC-C* tests. Consistency of identification is the percent overlap in identification of who the poorest performers are.

The first column of Table 3 shows the percent overlap when assessing the low-tail performers across the full sample. As indicated in the bottom row, only about half the time (54%) does a comprehender who performs poorly on one type of test also perform poorly on another type. The greatest similarity in identification (59%) is between reading comprehension tests that are *RC-C* and listening comprehension. The least similarity (48%) is between *RC-D* tests and listening. The differences between types of reading comprehension tests in their consistency of identification with listening reflects the fact that *RC-D* tests are more likely to select poor comprehenders who are poor decoders than either *RC-C* tests or listening comprehension tests.

Columns two and three of Table 3 show the consistency of low-tail identification broken down by age group. Using a median split (median = 11.17 years), we divided the sample into a younger group ($M=9.5$, $SD=.91$) and an older group ($M=13.89$, $SD=1.9$). As with the full sample, we again see that the least similarity in identification occurs between listening and *RC-D* tests.

What is most interesting from the age analysis is the clear trend for consistency in identification across modality to increase with the age of the comprehender. Comparing columns two and three shows that the overlap between reading and listening in who is identified as a poor comprehender is greater for the older age group. That holds for each type of reading test as well as the composite. The average increase is 10%. Thus, we learn from these data that the developmental shift that occurs in what accounts for reading comprehension performance as children age (Catts, et al., 2005; Curtis, 1980; Keenan, et al., 2008) has as one of its consequences that there is more consistency across modality of comprehension testing in the identification of poor performers.

4.2.3. Specifically Poor Comprehenders – Prevalence—Specifically poor comprehenders are those performing in the low tail on comprehension who have decoding skills at least one standard deviation higher; they are thus comprehending poorly on reading comprehension tests for reasons beyond poor decoding. Because decoding skill and comprehension are correlated in the general population, with correlations higher for reading

comprehension than for listening comprehension (see Table 2), it is of interest to know how often these skills dissociate. Table 4 presents our prevalence rates as a function of the selection tests, age, and whether decoding is assessed by word reading or nonword reading.

Although prevalence rates vary as a function of selection methods, on average only 7.5% of the sample qualified as specifically poor comprehenders when the discrepancy was assessed with word reading, and only 10% when it was assessed with nonword reading (averaging across the two modality composites). Regardless of the age of the comprehender or the type of selection test, using nonword reading as the basis for the discrepancy between decoding and comprehension led to higher prevalence rates of specifically poor comprehenders. The prevalence is higher when using a discrepancy with nonword reading because of the differential involvement of vocabulary knowledge; using a discrepancy with word reading has the potential to exclude those who are poor comprehenders because of poor vocabulary.

Table 4 shows that prevalence rates clearly vary not only as a function of whether word or nonword reading is used for the discrepancy, but also as a function of the type of selection test and the age of the comprehender. For example, only 1% of the younger sample qualified as being specifically poor comprehenders when comprehension was assessed with RC-D tests, whereas 7% qualified when comprehension was assessed with RC-C tests, and 11% when comprehension was assessed with listening comprehension. Thus, there is no consistent prevalence rate for specifically poor comprehension; it depends on selection method. Age matters too for prevalence rates. But it only affects prevalence when using reading comprehension tests for selection; not when using listening. Table 4 shows that there are more specifically poor comprehenders defined by reading comprehension among the older than the younger children.

4.2.4. Specifically Poor Comprehenders – Consistency—Table 5 shows consistency of identification of specifically poor comprehenders across the different selection tests when their discrepancy with decoding is defined relative both to word reading and to nonword reading. The first section of the table shows the results over the full sample, and the next two sections show results for the younger and older age groups. The results are a striking contrast to when poor comprehenders were defined as simply performing in the low-tail. Whereas consistency in identifying the same poor comprehenders in the low-tail occurred about half the time, consistency in identifying specifically poor comprehenders across different selection methods is much lower. For example, when they are defined to have a discrepancy with word reading, the likelihood of identifying the same individuals using reading tests versus listening tests averages only 11% for *RC-D* tests and 24% for *RC-C* tests. The consistency is even lower if the individuals are young. The consistency improves by 12% if the poor comprehenders are defined to have a discrepancy with nonword as opposed to word reading. But even when using a nonword reading discrepancy, the odds of any two methods identifying the same individuals as specifically poor comprehenders is less than a third – a rather dismal rate of consistency.

4.2.5. Profile Characteristics as a Function of Selection Method—Inconsistencies across selection methods in who is identified as a poor comprehender, especially in who is identified as specifically poor comprehenders, raises the question of whether there are

cognitive differences between the different individuals selected by the different methods. Table 6 displays cognitive profiles of the poor comprehenders identified by each method. Shown are Verbal IQ, Performance IQ, and mean z-scores from a composite of working memory tests and a composite of ADHD symptom counts.

Table 6 is divided into three sections according to selection criteria: low-tail poor comprehenders, specifically poor comprehenders using discrepancy with word reading, and specifically poor comprehenders using discrepancy with nonword reading. Within each section are the different comprehension tests that were used for selection: *RC-D*, *RC-C*, and Listening.

As is evident from Table 6, mean VIQ and PIQ differ very little as a function of which tests were used for selection. However, IQ does differ as a function of selection criteria. Namely, specifically poor comprehenders have higher VIQs than those of low-tail poor comprehenders ($F(1,378) = 14.43, p < .001$)². Furthermore, while low-tail poor comprehenders have VIQs that are quite similar to their PIQs, specifically poor comprehenders have higher VIQs than PIQs, resulting in a significant interaction of selection method with IQ, $F(1,378) = 13.21, p < .001$. Thus, researchers selecting specifically poor comprehenders are selecting individuals that differ from low-tail poor comprehenders not only by having higher decoding skill but also higher VIQs.

While VIQ of the poor comprehender varies as a function of selection method, but not as a function of test, what we see for working memory is the opposite - that it is the test that makes a difference, not the selection method. Specifically, poor comprehenders defined by *RC-D* tests have poorer working memory than those defined using the other tests. It does not matter whether poor comprehenders are low-tail or specifically poor, if they were selected with an *RC-D* test, then they have poorer working memory. For example, comparing working memory scores for the low-tail poor comprehenders selected by the three types of tests yielded a significant test effect, $F(2,1137) = 6.35, p < .005$, with post hoc *t*-tests (Games-Howell) showing those selected by *RC-D* tests to be significantly worse both compared to those selected by *RC-C* tests ($p < .01$) and compared to those selected by listening tests ($p < .01$). As we discuss later, this test difference in working memory can be attributed to the high memory load imposed by format features of these tests.

Poor comprehenders defined with a discrepancy in word reading average fewer ADHD symptoms than the other selection methods. However, ADHD symptom counts did not vary either by tests or by selection methods ($ps > .19$).

5. Discussion

Comprehension skills show continuous variations within the population. There is no break in the distribution separating poor comprehenders from others. Yet, for purposes of research on the nature and causes of poor comprehension, and for identifying children who need

²Because of the overlap in individuals between selection method conditions, tests of significance compared those uniquely identified by each selection method.

remediation, it is common to draw a line in the continuous distribution and call those performing below it poor comprehenders.

Everyone recognizes that if the line is drawn at the 10th percentile rather than the 25th that there will be differences in severity. It is much less common for all the other differences that exist in the literature in defining poor comprehenders to be recognized as having an impact. The present findings, however, show that such variations across studies should not be considered insignificant because they can dramatically alter who is selected and what their deficits are.

We assessed the extent to which these differences in comprehension tests and selection criteria matter by giving the same children multiple tests of reading and listening comprehension and by varying the criteria for defining poor comprehension and examining whether a child identified as a poor comprehender by one method was similarly identified if different tests or different selection criteria were used. For example, does it matter if a researcher uses a reading versus a listening comprehension test, or if they use different types of reading comprehension tests, or if they focus on specifically poor comprehenders versus low-tail poor comprehenders?

Because reading comprehension tests have been reported to differ, sometimes dramatically, in the extent to which their variance is explained by decoding skill (Cutting & Scarborough, 2006; Francis et al., 2005; Keenan, et al, 2008; Nation & Snowling, 1997), and because that difference has been shown to result in inconsistencies across reading tests in who they classify as performing in the low-tail (Keenan & Meenan, 2014), the present study took that difference into account when examining consistency of identifying poor comprehenders. We classified reading comprehension tests as either *RC-D* or *RC-C* and used composites of the test types to have more stable measures.

The overall finding is that the rates of consistency across different tests and selection methods in identifying poor comprehenders are rather low. The highest consistency was observed when poor comprehenders were defined simply as those performing in the low-tail. However, even then only half the time (54%) did a comprehender who performed poorly on one type of reading or listening comprehension test perform poorly on another type. Furthermore, in contrast to what one might expect, consistency was greatest not across the two types of reading tests, but rather across modality between *RC-C* and listening comprehension tests, reflecting the differences in whether tests identify weaknesses in oral language skills or identify as poor comprehenders those who are poor decoders.

The lowest consistency was across definitions of specifically poor comprehenders. This is a disturbing finding because those are the types of poor comprehension that most studies use. It might be tempting to think that because specifically poor comprehenders' decoding is so much greater than their comprehension and because test differences (e.g., *RC-C* vs. *RC-D*, reading vs. listening) reflect differences in decoding, then test differences may not matter when studying specifically poor comprehenders. However, we found just the opposite. Consistency across tests was much lower for specifically poor comprehenders than for low-

tail poor comprehenders. It averaged a dismal 15% when discrepancy with decoding was defined with word reading and was higher, but still quite low, 27% with nonword reading.

Why might consistency rates be so low? One factor could be the reliability of the comprehension tests. The published reliability statistics for the tests, however, are quite high, ranging from .64 - .98. But it should be noted that when defining specifically poor comprehenders, even when reliabilities of the selection tests are high, the reliabilities of difference scores tend to be lower than those for individual measures. Another factor relevant for the very low consistencies in defining specifically poor comprehenders is the small number of children who qualify. As the prevalence rates in Table 4 show, specifically poor comprehenders ranged from 1% -13% of the sample; thus, they constituted somewhat smaller samples than when we examined the lowest 15% for poor comprehenders that were not specifically poor. When only 1% are specifically poor comprehenders defined by *RC-D* tests and only 7% defined by *RC-C* tests, the small percentages are likely to affect consistency. But note that we had a sample of 1522 children. Most studies have far fewer in their sample. What these rates suggest is that definitional choices can alter who qualifies as a poor comprehender and, therefore, there are likely similar inconsistencies across different studies.

We gained insight into just how different poor comprehenders defined by different methods could be from our profile analyses. We found that both the type of test used for comprehension assessment and the type of selection criteria for defining poor comprehenders made a difference in the poor comprehender's cognitive profile. When poor comprehenders were defined as being specifically poor and thus having a discrepancy with decoding, we found that they had higher VIQs than PIQs and higher VIQs than those poor comprehenders defined as simply being in the low-tail of the distribution. Higher verbal skill in specifically poor comprehenders thus appears to be a byproduct of requiring higher decoding skills; it is because verbal skills like vocabulary are involved in decoding, at least in English. As would be expected then, the IQ difference was greater when the discrepancy was defined using word than nonword reading. Thus, even though it is common to think that specifically poor comprehenders differ from low-tail poor comprehenders only in decoding, we have learned now that there are also differences in VIQ.

The field has become quite cognizant of differences between reading tests in how much decoding explains their variance (e.g., Cutting & Scarborough, 2006; Francis et al., 2005; Keenan, et al, 2008; Nation & Snowling, 1997). But our profile analyses showed that there are other differences between tests that impact the type of person identified to be a poor comprehender. Specifically, *RC-D* tests selected as poor comprehenders those individuals with significantly poorer working memory than other poor comprehenders. This is an important difference between tests because working memory differences can impact other differences associated with poor comprehension, such as inferencing (Cain & Oakhill, 2006; Eason & Cutting, 2009).

It is ironic that *RC-D* tests would select for poor working memory more than the other tests when they have the shortest passages. However, when one considers format features of the tests and how the tests are administered, it becomes readily apparent that they impose

substantial working memory loads. The PIAT requires that each single-sentence text be held in memory as the child looks at four pictures to choose which picture best represents the meaning of the sentence. Similarly, the WJPC with its cloze format requires that each passage blank be held in memory while reading the rest of the passage and considering various word choices to fill in the blank.

In sum, we have shown that the different tests and different selection criteria that people use to define poor comprehenders can affect not only which individuals are selected as poor comprehenders but also the types of deficits that these individuals have. Practically speaking, the inconsistencies we observed across selection methods present a challenge for diagnosis; they suggest that the wiser course may be to withhold assignment of the poor comprehender label until grounded in multiple assessments, instead of performance from one test. At the same time, they provide important insights into the nature of comprehension and what it means to be a poor comprehender. We know now that different methods for defining poor comprehension select different individuals – individuals that sometimes can vary in the types of deficits they exhibit. Based on these findings, it seems imperative that we make the methods of selection of poor comprehenders in our research not only transparent in all our reporting, but also central to the interpretation of our findings.

Acknowledgments

This research was supported by a grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Grant No. P50 HD27802) to the Colorado Learning Disabilities Research Center. We are grateful to all the participants and their families, and all the testers and scorers.

References

- Barnes MA, Dennis M, Haefele-Kalvaitis J. The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from six to fifteen years of age. *Journal of Experimental Child Psychology*. 1996; 61:216–241.10.1006/jecp.1996.0015 [PubMed: 8636665]
- Betjemann RS, Keenan JM, Olson RK, DeFries JR. Choice of reading comprehension test influences the outcomes of genetic analyses. *Scientific Studies of Reading*. 2011; 15:363–382.10.1080/10888438.2010.493965 [PubMed: 21804757]
- Bianco M, Pellenq C, Lambert E, Bressoux P, Lima L, Doyen A. Impact of early code skill and oral comprehension training on reading achievement in first grade. *Journal of Research In Reading*. 2012; 35:427–455.10.1111/j.1467-9817.2010.01479.x
- Cain K, Oakhill J. Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*. 2006; 76:683–696.10.1348/000709905X67610 [PubMed: 17094880]
- Cain, K.; Oakhill, J. Reading comprehension difficulties: Correlates, causes, and consequences. In: Cain, K.; Oakhill, J., editors. *Children's comprehension problems in oral and written language*. New York: Guilford Press; 2007. p. 41-75.
- Cain K, Oakhill J, Bryant P. Investigating the causes of reading comprehension failure: The comprehension-age match design. *Reading and Writing: An Interdisciplinary Journal*. 2000; 12:31–40.10.1023/A:1008058319399
- Case R, Kurland M, Goldberg J. Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*. 1982; 33:386–404.
- Catts HW, Adlof SM, Weismer SE. Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language and Hearing research*. 2006; 49:278–293.10.1044/1092-4388(2006/023)

- Catts H, Hogan T, Fey M. Subgrouping poor readers on the basis of individual differences in reading related abilities. *Journal of Learning Disabilities*. 2003; 36:151–164.10.1177/002221940303600208
- Christopher ME, Miyake A, Keenan JM, Pennington BF, DeFries JC, Wadsworth SJ, Willcutt E, Olson RK. Predicting word reading and comprehension with executive function and speed measures across development: A latent variable analysis. *Journal of Experimental Psychology: General*. 2012; 141:470–488.10.1037/a0027375 [PubMed: 22352396]
- Clarke PJ, Snowling MJ, Truelove E, Hulme C. Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science*. 2010; 21:1106–1116.10.1177/0956797610375449 [PubMed: 20585051]
- Curtis ME. Development of components of reading skill. *Journal of Educational Psychology*. 1980; 72:656–669.
- Cutting L, Scarborough H. Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*. 2006; 10(3):277–299.10.1207/s1532799xssr1003_5
- Daneman M, Carpenter PA. Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*. 1980; 19:450–466.
- Defries JC, Filipek PA, Fulker DW, Olson RK, Pennington BF, Smith SD. Colorado Learning Disabilities Research Center. *Learning Disabilities*. 1997; 8:7–19.
- Dunn, LM.; Markwardt, FC. Examiner's manual: Peabody individual achievement test. Circle Pines, MN: American Guidance Service; 1970.
- DuPaul GJ, Power TJ, McGoe KE, Ikeda MJ, Anastopoulos AD. Reliability and validity of parent and teacher ratings of attention-deficit/hyperactivity disorder symptoms. *Journal of Psychoeducational Assessment*. 1998; 16(1):55–68.10.1177/073428299801600104
- Eason, SH.; Cutting, LE. Examining sources of poor comprehension in older poor readers: Preliminary findings, issues, and challenges. In: Wagner, RK.; Schatschneider, C.; Phythian-Sence, C., editors. *Beyond decoding: The behavioral and biological foundations of reading comprehension*. New York: Guilford Press; 2009. p. 263-283.
- Elwér Å, Keenan JM, Olson RK, Byrne B, Samuelsson S. Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of Experimental Child Psychology*. 2013; 115:497–516.10.1016/j.jecp.2012.12.001 [PubMed: 23528975]
- Francis, DJ.; Fletcher, JM.; Catts, HW.; Tomblin, J. Dimensions affecting the assessment of reading comprehension. In: Paris, SG.; Stahl, SA.; Paris, SG.; Stahl, SA., editors. *Children's reading comprehension and assessment*. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers; 2005. p. 369-394.
- Fricke S, Bowyer-Crane C, Haley AJ, Hulme C, Snowling MJ. Efficacy of language intervention in the early years. *Journal of Child Psychology And Psychiatry*. 2013; 54(3):280–290.10.1111/jcpp.12010 [PubMed: 23176547]
- Harlaar N, Cutting L, Deater-Deckard K, DeThorne LS, Justice LM, Schatschneider C, Petrill SA. Predicting individual differences in reading comprehension: A twin study. *Annals of Dyslexia*. 2010; 60:265–288.10.1007/s11881-010-0044-7 [PubMed: 20814768]
- Hoover WA, Gough PB. The simple view of reading. *Reading & Writing*. 1990; 2:127–160.10.1007/BF00401799
- Hulme C, Snowling MJ. Children's reading comprehension difficulties: Nature, causes, and treatments. *Current Directions in Psychological Science*. 2011; 20:139–142.10.1177/0963721411408673
- Keenan J, Betjemann R, Olson R. Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*. 2008; 12:281–300.10.1080/10888430802132279
- Keenan J, Betjemann R, Wadsworth S, DeFries J, Olson R. Genetic and environmental influences on reading and listening comprehension. *Journal of Research in Reading*. 2006; 29:75–91.10.1111/j.1467-9817.2006.00293.x
- Keenan JM, Meenan CE. Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities*. 2014; 47:125–135.10.1177/0022219412439326 [PubMed: 22442251]
- Keenan, JM.; Olson, RK.; Betjemann, RS. Assessment and etiology of individual differences in reading comprehension. In: Wagner, R.; Schatschneider, C.; Phythian-Sence, C., editors. *Beyond*

decoding: The biological and behavioral foundations of reading comprehension. New York: Guilford; 2009. p. 227-245.

Lahey BB, Applegate B, McBurnett K, Biederman J, Greenhill L, Hynd GW, Shaffer D. DSM-IV field trials for attention deficit hyperactivity disorder in children and adolescents. *American Journal of Psychiatry*. 1994; 151:1673–1685. [PubMed: 7943460]

Leslie, L.; Caldwell, J. *Qualitative Reading Inventory-3*. New York: Addison Wesley Longman; 2001.

Lieberman IY, Shankweiler D, Fisher FW, Carter B. Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*. 1974; 18:201–212.

Nation K, Cocksey J, Taylor J, Bishop D. A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry*. 2010; 51:1031–1039.10.1111/j.1469-7610.2010.02254.x [PubMed: 20456536]

Nation K, Snowling M. Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology*. 1997; 67(3):359–370. [PubMed: 9376312]

Nation K, Snowling M. Semantic processing and the development of word-recognition-skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*. 1998; 39:85–101.10.1006/jmla.1998.2564

Oakhill J, Cain K. The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*. 2012; 16:91–121.10.1080/10888438.2010.529219

Oakhill, J.; Cain, K. Introduction to comprehension development. In: Oakhill, J.; Cain, K., editors. *Children's comprehension problems in oral and written language*. New York: Guilford Press; 2007. p. 3-40.

Oakhill J, Cain K, Bryant PE. The dissociation of single-word reading and text comprehension skills. *Language and Cognitive Processes*. 2003; 18:443–468.10.1080/01690960344000008

Olson RK. Genes, environment, and dyslexia: The 2005 Norman Geschwind memorial lecture. *Annals of Dyslexia*. 2006; 56(2):205–238. [PubMed: 17849199]

Olson, R.; Forsberg, H.; Wise, B.; Rack, J. Measurement of word recognition, orthographic, and phonological skills. In: Lyon, GR., editor. *Frames of reference for the assessment of learning disabilities: New views on measurement issues*. 1994. p. 243-277.

Ouellette GP. What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*. 2006; 98:554–566.10.1037/0022-0663.98.3.554

Perfetti, CA. *Reading ability*. New York: Oxford University Press; 1985.

Pimperton H, Nation K. Suppressing irrelevant information from working memory: Evidence for domain-specific deficits in poor comprehenders. *Journal of Memory and Language*. 2010; 62:380–391.10.1016/j.jml.2010.02.005

Shankweiler, D. How problems of comprehension are related to difficulties in decoding. In: Shankweiler, D.; Liberman, IV., editors. *Phonology and Reading Disability: Solving the reading puzzle*. Ann Arbor: University of Michigan Press; 1989.

Sesma H, Mahone E, Levine T, Eason S, Cutting L. The contribution of executive skills to reading comprehension. *Child Neuropsychology*. 2009; 15(3):232–246.10.1080/09297040802220029 [PubMed: 18629674]

Spooner AR, Baddeley AD, Gathercole SE. Can reading accuracy and comprehension be separated in the Neale Analysis of Reading Ability? *British Journal of Educational Psychology*. 2004; 74:187–204.10.1348/000709904773839833 [PubMed: 15130187]

Stothard SE, Hulme C. Reading comprehension difficulties in children: The role of language comprehension and working memory skills. *Reading and Writing*. 1992; 4:245–256.10.1007/BF01027150

Swanson, H.; Howard, CB.; Sáez, L. Reading comprehension and working memory in children with learning disabilities in reading. In: Cain, K.; Oakhill, J., editors. *Children's comprehension problems in oral and written language: A cognitive perspective*. New York, NY US: Guilford Press; 2007. p. 157-185.

- Vellutino FR, Tunmer WE, Jaccard JJ, Chen R. Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*. 2007; 11:3–32.10.1207/s1532799xssr1101_2
- Wechsler, D. Manual for the Wechsler intelligence scale for children, revised edition. San Antonio, TX: The Psychological Corporation; 1974.
- Wechsler, D. Manual for the Wechsler adult intelligence scale, revised edition. San Antonio, TX: The Psychological Corporation; 1981.
- Wiederholt, L.; Bryant, B. Examiner's manual: Gray Oral Reading Test-3. Austin, TX: Pro-Ed; 1992.
- Willcutt E, Pennington B, Olson R, Chhabildas N, Hulslander J. Neuropsychological analyses of comorbidity between reading disability and attention deficit hyperactivity disorder: In search of the common deficit. *Developmental Neuropsychology*. 2005; 27(1):35–78.10.1207/s15326942dn2701_3 [PubMed: 15737942]
- Woodcock RW, McGrew KS, Mather N. Woodcock-Johnson III tests of Achievement. 2001
- Yuill, N.; Oakhill, J. Children's problems in text comprehension. Cambridge: Cambridge University Press; 1991.

Table 1
Standard Scores for Measures Used in the Study That Are Nationally Normed

Measure	Mean	Standard Deviation
Full-Scale IQ	107.65	13.72
Verbal IQ	109.68	14.87
Performance IQ	104.06	13.46
Reading Comprehension Tests		
GORT-3	11.21	3.22
PIAT	106.28	12.84
WJPC-3	101.88	10.43
Word Decoding		
PIAT Test of Word Reading	104.52	12.34

Table 2
Correlations among Variables Used for Identifying Poor Comprehenders

Measure	Reading Composite	RC-C	RC-D	Listening Composite	Word Reading	Nonword Reading
Reading Composite	1.0					
RC-C	.91	1.0				
RC-D	.92	.68	1.0			
Listening Composite	.74	.71	.64	1.0		
Word Reading	.73	.54	.78	.48	1.0	
Nonword Reading	.59	.42	.65	.35	.83	1.0

* All values are significant at $p < .001$.

Table 3
Consistency of Identification (% Same Children) of Poor Comprehenders (Lowest 15%)
across Different Comprehension Tests for Full Sample, and for Younger and Older Age
Groups

Test Comparison	Full Sample	Younger Mean = 9.5 yrs	Older Mean = 13.89 yrs
Read Composite Listen Composite	57	52	65
Read RC-D Listen Composite	48	45	54
Read RC-C Listen Composite	59	53	62
Read RC-D Read RC-C	53	55	52
Mean Consistency	54	51	58

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4
Prevalence of Specifically Poor Comprehenders (Lowest 25% with at least 1SD Discrepancy with Decoding) As a Function of Age and Whether Decoding is Defined by Word Reading or Nonword Reading

Comprehension Test	Full Sample		Young Mean = 9.5 yrs		Older Mean = 13.9 yrs	
	Word	Nonword	Word	Nonword	Word	Nonword
Reading Comprehension Composite	5%	8%	3%	6%	6%	10%
RC-D Composite	2.5%	6%	1%	4%	4%	9%
RC-C Composite	8%	11%	7%	9%	9%	13%
Listening Comprehension Composite	10%	12%	11%	13%	9%	11%

Table 5
Consistency of Identification (% Same Children) of Specifically Poor Comprehenders
(Lowest 25% with at least 1SD Discrepancy with Decoding) across Different
Comprehension Tests When Decoding is Assessed with either Word Reading or Nonword
Reading

Test Comparison	Discrepancy with Word Reading	Discrepancy with Nonword Reading
Full Sample		
Read RC-D Listen Composite	11	23
Read RC-C Listen Composite	24	36
Read RC-D Read RC-C	10	23
Younger Half of Sample		
Read RC-D Listen Composite	6	15
Read RC-C Listen Composite	19	33
Read RC-D Read RC-C	7	19
Older Half of Sample		
Read RC-D Listen Composite	17	31
Read RC-C Listen Composite	33	41
Read RC-D Read RC-C	11	25

Table 6
Mean Verbal IQ and Performance IQ standard scores and Mean z-scores (standard deviation) for Working Memory and ADHD Symptom Count for Each Selection Method

Selection Method	VIQ	PIQ	Working Memory	ADHD Symptoms
Low-tail				
RC-D	95 (12)	96 (13)	-.70 (.75)	.24 (1.07)
RC-C	96 (12)	95 (12)	-.52 (.84)	.21 (1.09)
Listening	96 (12)	96 (13)	-.51 (.88)	.21 (1.13)
Low-tail + Discrepancy with Word Reading				
RC-D	98 (11)	94 (14)	-.50 (.78)	.03 (1.03)
RC-C	101 (13)	96 (12)	-.20 (.85)	.10 (1.06)
Listening	100 (12)	96 (13)	-.21 (.89)	.16 (1.15)
Low-tail + Discrepancy with Nonword Reading				
RC-D	95 (11)	92 (13)	-.65 (.79)	.24 (1.03)
RC-C	98 (12)	95 (12)	-.34 (.86)	.21 (1.12)
Listening	98 (12)	95 (13)	-.34 (.87)	.21 (1.14)