The Teacher Rating of Oral Language and Literacy (TROLL) is an instrument that measures skills identified as critical in the New Standards for Speaking and Listening. In 5 to 10 minutes and without prior training, teachers can assess an individual child's current standing with respect to skills that research has identified as critical for literacy acquisition. Skills assessed include language, reading, and writing abilities. TROLL has been used with over 900 low-income children. The instrument is reliable and has strong internal consistency. Its validity has been established in numerous ways; TROLL correlates significantly with scores on the Peabody Picture Vocabulary Test and the Early Phonemic Awareness Profile given to the same children by trained researchers. Fall TROLL judgments of the literacy skills of these same children correlate with fall but not spring assessments of literacy skills obtained with the Early Literacy Profile. TROLL also displays instructional sensitivity. An appendix contains "Teacher Rating of Oral Language and Literacy" (Education Development Center, 1997). (Contains 3 tables and 24 references.) (PM)
Teacher Rating of Oral Language and Literacy (TROLL): A Research-Based Tool

David K. Dickinson, Educational Development Center
Allyssa McCabe, University of Massachusetts—Lowell
Kim Sprague, Educational Development Center
Teacher Rating of Oral Language and Literacy (TROLL): A Research-Based Tool

CIERA REPORT #3-016

David K. Dickinson, Educational Development Center
Allyssa McCabe, University of Massachusetts—Lowell
Kim Sprague, Educational Development Center

CIERA Inquiry 3: Policy and Profession

What is the Teacher Rating of Oral Language and Literacy (TROLL)?
Can it provide an accurate measure of students’ literacy acquisition skills?

In this article, we describe the Teacher Rating of Oral Language and Literacy (TROLL), an instrument that measures skills identified as critical in the New Standards for Speaking and Listening (Tucker & Codding, 1998). In 5–10 minutes and without prior training, teachers can assess an individual child’s current standing with respect to skills that research has identified as critical for literacy acquisition. Skills assessed include language, reading, and writing abilities. TROLL has been used with over 900 low-income children. The instrument is reliable and has strong internal consistency. Its validity has been established in numerous ways; TROLL correlates significantly with scores on the Peabody Picture Vocabulary Test and the Early Phonemic Awareness Profile given to the same children by trained researchers. Fall TROLL judgements of the literacy skills of these same children correlate with fall but not spring assessments of literacy skills obtained with the Early Literacy Profile. TROLL also displays instructional sensitivity.

Center for the Improvement of Early Reading Achievement
University of Michigan – Ann Arbor

September 20, 2001
©2001 Center for the Improvement of Early Reading Achievement.

This research was supported under the Educational Research and Development Centers Program, PR/Award Number R305R70004, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. However, the comments do not necessarily represent the positions or policies of the National Institute of Student Achievement, Curriculum, and Assessment or the National Institute on Early Childhood Development, or the U.S. Department of Education, and you should not assume endorsement by the Federal Government.
To be able to read and write, children first must develop sufficient oral language skills (Dickinson & McCabe, 1991; Dickinson & Tabors, 2001; Snow 1983; Snow, Burns, & Griffin, 1998). Oral language skills blossom during the preschool years, which means that they are also very vulnerable and in need of stimulation during this time, as a number of major organizations involved in the education of young children have recognized.

Because speaking and listening are so critical for literacy development in early childhood, the New Standards project, a program of the National Center on Education and the Economy (Tucker & Codding, 1998), has developed research-based standards for speaking and listening for preschool through third grade. These standards complement those already developed for reading and writing. Speaking and listening standards include many specific recommendations for teachers regarding beneficial habits of conversation, useful kinds of talk such as narratives and explanations, and language conventions relevant to early childhood. Many programs will be concerned to see that their students meet such national standards.

Anna is a quiet, attractive Hispanic girl, with long wavy black hair always tied back neatly into a ponytail and large brown eyes that seem to study everything around her. Her teacher has observed that she is shy with other children, reluctant to participate in group choruses, and often the last to join a line or in clean-up activities. Still, the teacher has never had to reprimand Anna. Only when the teacher sits down to fill out reports on the language development of each of her students does she realize Anna has a problem.

A Call for Developmentally Appropriate Assessment

In 1998, the International Reading Association (IRA) and the National Association for the Education of Young Children (NAEYC) joined together to formulate a position statement regarding early literacy development. The
statement acknowledges the difficulty that teachers face, for example, in kindergarten classrooms where a five-year range in children's literacy-related skills is not uncommon (Riley, 1996). Estimating where each child is in terms of the acquisition of speaking, listening, reading, and writing skills is critical in order to provide developmentally appropriate instruction to all children in this wide range. The position statement (p.38, emphasis ours) is quite clear that throughout these critical years accurate assessment of children's knowledge, skills, and dispositions in reading and writing will help teachers better match instruction with how and what children are learning. However, early reading and writing cannot simply be measured as a set of narrowly defined skills on standardized tests. These measures often are not reliable or valid indicators of what children can do in typical practice, nor are they sensitive to language variation, culture, or the experiences of young children.

As if these difficulties were not enough, teachers typically have not been trained to evaluate children's language development as it relates to the acquisition of literacy. In response to this need, this report presents an accessible means of evaluating each child in your classroom for literacy-related language development. Periodic evaluation of children using this tool will enable teachers and programs to assess and demonstrate their effectiveness.

**The Development of TROLL**

The Teacher Rating of Oral Language and Literacy (TROLL) is a rating tool developed by David Dickinson to provide teachers with a way to track the language and literacy development of children in their classrooms. It is represented in its entirety in the Appendix. TROLL is designed to enable teachers to rate a child's literacy-related skill and describe his or her literacy-related interests. Although it was developed for research purposes separate from the New Standards initiative, TROLL addresses all the central speaking and listening skills present in the New Standards. In fact, in constructing TROLL, the present researchers (who were also involved in developing these standards) drew on the same body of theory and research that was the foundation for *Speaking and Listening from Preschool through Third Grade*. TROLL also covers many of the early reading and writing skills teachers will want to track. Furthermore, TROLL allows teachers to track children's interest in various language and literacy activities—something that no direct assessment tool can hope to capture.

**Using TROLL**

No formal training is required for using TROLL. This tool is designed so that classroom teachers can easily track the language and literacy development of
TROLL only requires approximately 5-10 minutes for each child and, with a little planning, can be completed without disrupting classroom activities.

Teachers can use the information to inform their teaching. First, teachers can identify children who are most in need of rich language and literacy stimulation in order to catch up to their peers in this regard. Second, teachers could combine the results for all children in their classes to determine which areas of performance might merit more systematic instruction for the entire class, or at least a group of children. For example, if all children in a class score relatively low on rhymes, their teacher might want to begin providing numerous opportunities to listen to and produce rhyming chants, songs, and poems. Or, more likely, a teacher might find that half or two thirds of the class could benefit from more rhyming activity.

School reading specialists can rate children in a classroom independently of the classroom teacher. Comparing ratings of individual children can serve as a reliability check and as an excellent, focused springboard for discussion. Furthermore, the reading specialist might consider using TROLL ratings of children as a grounding for discussion with parents.

Parents constitute a third group of potential TROLL users. Teachers of bilingual children often have a very difficult time rating the language competence of children who speak English as a Second Language. In such cases, educators are advised to enlist the help of these children's parents. In fact, we found that maternal reports of preschoolers' literacy (when children were three or four years old) significantly predicted much of the variation in kindergarten tests, first-grade teacher assessments, and direct assessments of decoding given near the end of first grade (Dickinson, & DeTemple, 1998).

---

**The Contents of TROLL**

TROLL contains three subscales: (a) language use, (b) reading, and (c) writing. Introductory questions determine the language(s) the child speaks and his or her comprehension and production abilities in English. Teachers are given the opportunity to rate English and native language competence. The tool has 25 items each measured on a scale that varies slightly by item (i.e., most but not all use a 4-point scale). Total scores are easily calculated simply by adding individual scores on these 25 scales; total scores vary from a minimum of 24 to a maximum of 98. These total scores provide teachers both an indication of an individual child's development relative to other children and a means to chart an individual child's growth.

In Table 1 we display what different scores on TROLL indicate about a child's overall developmental level. For example, a score of 61 indicates that the child is making progress that is average for four-year-olds in the fall. We have reported the TROLL total score that corresponds to particular percentiles by converting the raw scores of our total sample to percentiles. This norming sample was composed of low-income, high-risk children, so that these norms should be regarded as provisional. However, we argue that the well-known academic disadvantages of low socioeconomic status (SES) pre-
school children (e.g., Stipek & Ryan, 1997) make this norming sample important in its own right, especially for those who work with low-income children. If a child from a low SES family scores at the 10th percentile, for example, this result cannot be dismissed as simply due to economic disadvantage; such a child is scoring very poorly relative to his or her economic peers.

Table 1: What TROLL Scores Mean

<table>
<thead>
<tr>
<th>3 year olds' TROLL scores</th>
<th>4 year olds' TROLL scores</th>
<th>5 year olds' TROLL scores</th>
<th>Relative standing on the TROLL</th>
<th>Recommendations/Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall, n=115</td>
<td>Spring, n=55</td>
<td>Fall, n=336</td>
<td>Spring, n=234</td>
<td>Fall, n=83</td>
</tr>
<tr>
<td>40</td>
<td>44</td>
<td>43</td>
<td>46</td>
<td>51</td>
</tr>
<tr>
<td>44</td>
<td>49</td>
<td>52</td>
<td>55</td>
<td>59</td>
</tr>
<tr>
<td>51</td>
<td>56</td>
<td>61</td>
<td>66</td>
<td>68</td>
</tr>
<tr>
<td>61</td>
<td>62</td>
<td>71</td>
<td>74</td>
<td>75</td>
</tr>
<tr>
<td>68</td>
<td>69</td>
<td>80</td>
<td>84</td>
<td>85</td>
</tr>
</tbody>
</table>

Psychometric Properties of TROLL

Field Testing

More than 900 children have been assessed over the last several years. New England Quality Research Center for Head Start (NEQRC) used this instrument, as did the Literacy Environment Enrichment Program (LEEP), a professional development program designed for use with preschool teachers and their supervisors.1

Reliability

Teachers independently rated 272 children twice, and their ratings were comparable. Children who scored high on TROLL the first time also scored high the second time they were assessed, despite the fact that teachers did not have access to their prior ratings.2
This tool has high internal consistency, with alphas exceeding .89 for all ages (see Table 2). In other words, there is good reason to believe that the instrument is measuring a child's general oral language acumen rather than separate oral language skills. Note that alphas for separate language, reading, and writing subscales by age ranged from .77 to .92, representing strong internal consistency.

Table 2: Descriptive Statistics and Internal Consistency Measures

<table>
<thead>
<tr>
<th>AGE</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>RANGE</th>
<th>ALPHA*</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 years Fall</td>
<td>115</td>
<td>52.64</td>
<td>11.89</td>
<td>29-87</td>
<td>.94</td>
</tr>
<tr>
<td>Spr</td>
<td>55</td>
<td>56.40</td>
<td>10.19</td>
<td>36-81</td>
<td>.89</td>
</tr>
<tr>
<td>4 years Fall</td>
<td>336</td>
<td>61.63</td>
<td>13.61</td>
<td>30-97</td>
<td>.95</td>
</tr>
<tr>
<td>Spr</td>
<td>234</td>
<td>65.13</td>
<td>13.85</td>
<td>29-95</td>
<td>.95</td>
</tr>
<tr>
<td>5 years Fall</td>
<td>83</td>
<td>67.12</td>
<td>13.62</td>
<td>29-96</td>
<td>.94</td>
</tr>
<tr>
<td>Spr</td>
<td>229</td>
<td>74.21</td>
<td>13.78</td>
<td>37-97</td>
<td>.95</td>
</tr>
<tr>
<td>All Ages Fall</td>
<td>534</td>
<td>60.55</td>
<td>14.00</td>
<td>29-97</td>
<td>.95</td>
</tr>
<tr>
<td>All Ages Spr</td>
<td>518</td>
<td>68.22</td>
<td>14.70</td>
<td>29-97</td>
<td>.95</td>
</tr>
</tbody>
</table>

*Cronbach's alpha (Cronbach, 1951).

Validity

How well does this tool actually measure what it is intended to measure? This tool relies on a teacher's professional judgement or perception of a child's development rather than formal testing of actual development. It is therefore reassuring to find that the child ratings teachers provide using TROLL are largely consistent with those obtained by formal testing. Researchers administered the following formal tests to children who had been independently rated by their teachers using TROLL. The range of correlations between TROLL and the tests is reported in Table 3.

1. The well-established Peabody Picture Vocabulary Test (PPVT-III; Dunn & Dunn, 1997) is a measure of receptive vocabulary. We used the PPVT raw scores because TROLL itself is a raw score.

2. The Emergent Literacy Profile (ELP; Dickinson & Chaney, 1997) assesses children's ability to read environmental print (e.g., picks out "McDonald's"® from the logo) as well as children's sense of printed language (e.g., picks the word out of a display such as: NNNT, W3#NJ, MILK), their ability to identify letters, and their early attempts at writing. This tool was developed by the New England Research Center on Head Start Quality. Cronbach's alpha, a measure of internal consistency, equals .86 (n=578), representing strong consistency.

3. The Early Phonemic Awareness Profile (EPAP; Dickinson & Chaney, 1997) assesses children's ability to engage in phoneme deletion (e.g., what is "foot" without the "f"?) and recognize rhymes. Like the ELP, this tool was developed by the New England Research Center on Head Start Quality. Cronbach's alpha equals .93 (n=563), representing very strong consistency.
Table 3: Correlations* Between TROLL Subscales and Other Measures of Language and Literacy Development—QRC and LEEP data—All Ages Combined (n=461-688).†

<table>
<thead>
<tr>
<th></th>
<th>TROLL—Total Score‡</th>
<th>TROLL—Language n=304/287</th>
<th>TROLL—Reading n=295/275</th>
<th>TROLL—Writing n=306/278</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FALL</td>
<td>SPRING</td>
<td>FALL</td>
<td>SPRING</td>
</tr>
<tr>
<td>PPVT—Raw Score n=131/131</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.47***</td>
<td>.42***</td>
<td>.47***</td>
<td>.40***</td>
</tr>
<tr>
<td>Emergent Literacy Profile—Total n=68/67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.43**</td>
<td>ns</td>
<td>.42**</td>
<td>ns</td>
</tr>
<tr>
<td>Early Phonemic Awareness Profile—Total n=157/133</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.45***</td>
<td>.47***</td>
<td>.34***</td>
<td>.46***</td>
</tr>
</tbody>
</table>

†The EPAP and PPVT were administered in English to all children, regardless of language dominance. The ELP was administered in the child’s dominant language.
‡TROLL total scores fall and spring r=.88***.
* p < .05
** p < .01
*** p < .001

Teacher ratings of children’s language and literacy development on TROLL show moderate associations with children’s scores on all three formal measures. In other words, in about five minutes and with no special training, teachers themselves can index what specially trained researchers would spend 25–30 minutes per child assessing. Of course, the TROLL teacher ratings do not agree completely with the researchers’ tests. This difference partly reflects the fact that TROLL captures other factors that teachers take into consideration as they rate individual children—factors that are not captured in researchers’ direct, formal assessments. In other words, TROLL captures the kind of assessment information recommended in the position statement by IRA and NAEYC. Formal tests measure how well a child does at only one point in time; children may be tired or sick on the day of the PPVT-III assessment and receive a dismal score for their receptive vocabulary, whereas their teacher knows that on most days they are quick to pick up on the vocabulary of classroom units and articulate when sharing stories of personal experience. The TROLL score is not as vulnerable to fluctuations in a child’s performance as the formal tests. Also, as noted earlier, TROLL includes information about children’s engagement in literacy activities and patterns of oral language use.

TROLL scores correlate well with our measure of phonological awareness (EPAP), as is shown in Table 4. Numerous researchers have established that phonological awareness is a critical precursor, correlate, and predictor of reading achievement (Bryant, MacLean, & Bradley, 1990; Cronin & Carver, 1998; Speece, Roth, Cooper, & de la Paz, 1999; Stanovich, 1992; Vellutino & Scanlon, 2001; Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993; Wagner et al., 1997).

Another point worth noting is that TROLL total scores are more closely related to formal language scores—phonological awareness and vocabulary—than they are to literacy scores (see Table 4). Of special significance is the fact that, in the fall, teacher’s ratings of children’s print-related knowl-
edge correlate with children's scores on the Early Literacy Profile (e.g., environmental print, letter identification) whereas no such relationship exists in the spring—the opposite of what one would predict.

How can this be? After all, teachers have had far more interaction with children—and should therefore be more sensitive—in the spring than in the fall. One possibility is that in these preschool classrooms, throughout the year teachers have more occasions to observe children's spoken language than their literacy activities. In the fall they may make special efforts to identify children's level of literacy development, but either fail to revisit this assessment or have relatively few occasion to observe children using literacy so have no basis for accurately gauging their growth. Thus, children may make progress in their knowledge of print as registered by the ELP that goes undetected by their teachers. Alternatively, children may not be making such progress and teachers fail to note this. This finding of reduced ability to gauge children's literacy status is sobering. The danger of any teacher judgment is that it become a self-fulfilling prophecy (Rosenthal & Jacobsen, 1968); teachers' opinions of children at the very outset of their education may predict children's success. They may give extra attention and motivation to the children they expect to become the most accomplished; conversely, they may fail to support children they expect to fail.

This lack of a revision of judgements about individual children's literacy skills and interests is regrettable and points to the potential value of teachers carrying out periodic informal assessments of children to provide concrete evidence of children's growth. For example, teachers can ask children to identify letters in their own names and in those of other children in the classroom, to write their names, etc.

In short, rather than promote such self-fulfilling prophecies, we hope that TROLL will predict possible failure in order to prevent failure. That is, we hope that teachers will use this instrument to give children who are struggling the help they need so that they never have to experience true failure at a later point.

TROLL Measures Effective Instruction

Teachers often struggle to keep track of children's growth. Researchers have asked teachers to track over 900 children using TROLL. In the New England Quality Research Center study of Head Start programs, there was a significant improvement in 272 children's TROLL scores in the spring versus the fall ($t_{271}=14.65$, $p=.0001$). Teachers were developing children's language and literacy skills, and TROLL reflected the gains made by individual children. In other words, TROLL displays instructional sensitivity.

TROLL has also been able to detect changes that occurred as a result of a program improvement effort. Head Start teachers and their supervisors volunteered to participate in the LEEP program (previously introduced). They received academic credit for participating in two intensive three-day blocks three months apart. TROLL scores for children whose teachers participated in LEEP were significantly higher than for children whose teachers did not.
Specifically, children in LEEP classrooms gained more overall from fall to spring on average, compared to a control group (F=9.24, p=.003).

Furthermore, the classrooms that supported such advances were independently observed to have improved classroom language and literacy practices. Teachers who participated in LEEP made a significantly greater effort to engage children in conversation and to provide opportunities for children to write and use books. The most major shift was represented by the extent to which teachers planned activities with the intention of having children practice literacy-related skills. Enriching the literacy environment had one additional surprising impact: Children whose teachers participated in LEEP displayed significant growth in social skills as compared with their peers (F=8.46, p=.004). Children who are busy talking, reading, and writing—activities registered by TROLL—were more likely to be viewed by their teachers as developing stronger collaboration skills.

A Final Note

We conclude by returning to Anna's story. Anna scored the lowest on TROLL of anyone in her class—a total of 40 points. Her teacher realized that she spent far more time talking to Anna's high-scoring classmates than she did to Anna. Children who were already the most advanced talkers were the ones who asked her for the names of things that interested them, participated in group discussions, and took many opportunities to explain activities to other children or tell stories about themselves. In other words, the teacher found that in her classroom, as in the classrooms of other preschool teachers, the old adage "The rich get richer" applied to language and literacy development. The teacher also realized that several of the children who acted out frequently also received low scores, and she made an effort to involve them in more conversations. Two such boys, in particular, seemed to thrive on this extra attention for positive behavior, and were noticeably better behaved by the end of the year.

In the subsequent weeks, Anna's teacher made a concentrated effort to involve her in talk every day. Anna looked at her politely during these initiatives, but she seldom responded. Anna was still very, very quiet. On her teacher's recommendation, Anna's parents had her hearing acuity evaluated by an audiologist, who discovered that Anna had a mild hearing loss. Anna, like many other children (Friel-Patti & Finitzo, 1990), had previously undetected hearing loss that made it hard for her to learn in class. With medical treatment, Anna's hearing problem was readily resolved. However, she still struggled with language learning and delay and benefited from speech and language remedial services. Because her hearing loss was discovered so early, her parents, teacher, and speech language pathologist believe she has every chance to speak, read, and write with her peers. She did not have to fail at reading in order to get the help she needed to succeed. Anna benefited from the kind of early intervention strongly recommended by Snow, Burns, and Griffin (1998, pp. 318-319).
References


Notes

1. For more information about LEEP contact David Dickinson at EDC, 55 Chapel St., Newton, MA 02458.

2. It was not feasible to obtain test-retest reliability estimates, as teachers would have had to rate the same children twice within a short amount of time, a request that would have been unreasonable. Also, a close retest might have allowed examiner memory (and hence bias) to compromise the second rating.

3. There were several small changes in the technical details of the TROLL making it inappropriate to conduct psychometric assessments using the sample of 973. Hence we reduced the sample size. Analyses with the earlier sample had similar patterns.

4. This finding came from teacher ratings using an adaptation of another tool, the Social Skills Rating System (Gresham & Elliott, 1990).

5. See the report of patterns of classroom interaction in preschool classrooms in Dickinson and Tabors (2001).
NOTICE

Reproduction Basis

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

X This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").