Validity of the Early Years Evaluation

The Learning Bar[1]

I. Introduction

The *Early Years Evaluation* (*EYE*) is an evaluation system that provides leading indicators of children's early development. It comprises three complementary assessments that can be used together or separately to meet the needs of teachers and educational administrators. They assess children's skills in nine domains, consistent with the frameworks established by UNICEF and the US National Reading Panel.[2, 3] The skills are closely related to children's literacy development and their success at school. [4, 5, 6]

---

[1] This report was prepared by Dr. J. Douglas Willms with the support of the staff at The Learning Bar Inc.

[2] Shepard, L., Kagan, S. L., & Wurtz, E. (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: National Education Goals Panel.

[3] National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). National Institute of Child Health & Development.

[4] Torgesen, J. K. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. *American Educator, 22*, 32-39.

[5] Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia, 48*, 115-136.

[6] Hoover, W., & Gough, P. (1990). The simple view of reading. *Reading and Writing, 2*, 127-160.

Three Complementary Assessments

EYE-Direct Assessment (EYE-DA)

The *EYE-DA* is designed to measure the developmental outcomes of children aged 3-5 as they prepare for and make the transition to school. It is individually administered by a classroom teacher or trained evaluator and requires approximately 45 minutes to complete. It assesses skills in four domains.

Awareness of Self and Environment refers to the child's knowledge of the world and how it works. It includes concepts learned at home, in the community, and at school. This domain is important because it is the lens through which future information is filtered. It influences what children notice and think about and how they organize new information gleaned from future learning experiences.

Cognitive Skills are *thinking* skills. They include a range of skills associated with a child's ability to observe, concentrate, ask questions, and solve problems. Cognitive skills are the foundation for learning in a wide range of domains, such as learning a language, making art and music, creative expression, and solving mathematical and scientific problems. The cognitive skills for this domain include foundational pre-literacy and numeracy skills such as naming letters, sound-letter relationships, counting, and matching numbers with sets of objects.

Language and Communication Skills include *receptive language skills* – the ability to listen to and understand instructions, discussions, and stories – and *expressive language skills* – the ability to communicate and be understood by others. Children with good receptive language can better understand the language of school activities, follow classroom instructions, participate in discussions, and learn to read. Children who communicate well also find it easier to form relationships with teachers and classmates.

Physical Development Skills include *fine motor skills* – the ability to use small movements in activities, such as picking up and manipulating objects, drawing, writing, and using a keyboard – and g*ross motor skills* – the ability to use large movements of the body in activities such as running, jumping, crawling and climbing. Children's motor skills, coordination, energy levels, and physical health affect their ability to successfully participate in many school experiences that foster early learning, such as art activities that require drawing and using scissors, and social activities such as playing games at recess.

The assessment is available in English and French.

EYE-Teacher Assessment (EYE-TA)

The *EYE-TA* is a skills-based assessment tool used by kindergarten teachers for children aged 4 to 6 as part of their regular classroom practice. It assesses skills in four domains of the *EYE-DA*, as well as students' social skills and approaches to learning.

> Social Skills and Approaches to Learning refer to a child's ability to get along with others, make friends, follow directions, and regulate emotional reactions, and their approaches to learning, such as a willingness to try new activities and maintain concentration while playing. Children who interact well with teachers and peers form strong social relationships and do better academically.

Teachers collect formative data on their students' developmental skills over four to six weeks, typically at the beginning of the school year. Then, they assess each student's skills on a set of items using a four-point rubric, following the approach recommended by Robert Marzano. A score of 3 indicates 'mastery' of a skill or learning objective. A score of 1 indicates that the skill is emerging or nonexistent, while a score of 2 indicates the student is learning the skill but has not yet mastered it. A score of 4 is reserved for performances that exceed expectations.[7]

The assessment is available in English and French.

[7] Marzano, R. J. (2010). *Formative assessment and standards-based grading*. Bloomington, IN: Marzano Research Laboratory.

EYE Pre-Reading Assessment (EYE-PR)

The *EYE-PR* is a short 10-item assessment that assesses pre-reading skills in four domains:

Letter Knowledge is the ability to recognize and name lower- and upper-case letters automatically.

Sound-letter Relationships refer to the association between the sounds we hear in words (phonemes) to the letters or combinations of letters (graphemes) we see in print. These are also called phoneme-grapheme relationships or PGRs.

Phonological Awareness is the ability to recognize and manipulate the phonological structures of spoken words. Two skills commonly associated with phonological awareness are the ability to segment words into syllables and blend syllables into words.

Phonemic Awareness is the ability to hear, identify, and manipulate the phonemes of spoken words. Four of the skills commonly taught in kindergarten and grade 1 include blending phonemes into syllables or words, segmenting syllables into phonemes or their onset and rime, isolating the initial and final parts of syllables and words, and deleting the initial and final phonemes of a word.

The *EYE-PR* is individually administered by a classroom teacher or trained evaluator and requires approximately 15 minutes to complete. The results inform teachers and educational administrators about the specific pre-reading skills children acquire that prepare them to learn how to read.

II. Validity

The early theoretical discussions of validity were concerned with *construct* validity – does a test measure what it purported to measure? The 1974 *Standards for Educational and Psychological Tests* recognized three types of validity – content, concurrent, and predictive. These concepts became the focus of research on test validity for several decades. For example, a test could be deemed 'valid' if the content adequately sampled the objectives of a curriculum, if the scores on the test were correlated with scores from other widely used tests, and if the scores were predictive of competencies or some event measured in the future.

The current approach to validity focuses on the use and interpretations of an assessment:

> "Validity refers to the degree to which evidence and theory support the interpretation of test scores for proposed uses of tests". [8]

Validity rests on the coherence and clarity of the argument that links test results to their intended interpretation and use.[9] Therefore, one cannot claim that a test is either valid or invalid; rather, one must ask whether the proposed interpretations and uses of the test results are valid. When parents, teachers, principals, or school administrators receive *EYE* results, how are they interpreted and used? What kinds of decisions are being made?

The research on the validity of the instruments developed by The Learning Bar is concerned with the theory and evidence that supports a 'validity claim' about a proposed use of a test. Some of the key questions include:

- Who collects the data? How is it collected and under what conditions?
- How are the test performances scored?
- Does the assessment cover a representative sample of the learning objectives of the curriculum? (Content validity)
- Are test scores correlated with scores based on previously validated tests and with measures? Are they predictive of future events? (Criterion-related validity)

---

[8] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

[9] Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1-73.

- Is the measurement process consistent across testing occasions, evaluators, and contexts? (Reliability)
- Can we add meaning to the scores by referencing them to mastery of specific competencies or to norms for different groups?

The theory and evidence supporting a validity claim are collectively referred to as the 'validity argument.' The validity argument backs up a claim that a test can be used for a specific purpose. For example, we consider the *EYE-PR* to be a valid assessment for screening purposes, while the *EYE-TA* is a valid assessment for informing teachers' daily practice. Together, they are valid for several uses.

The next section describes characteristics of assessments that can support claims about their validity. These apply to all educational assessments. The validity claims for the *EYE* are presented, along with the research, evidence, and resources that support them.

The final section of the paper presents the validity claims and the validity arguments for five uses of the *EYE*:

- improving classroom instruction
- allocating resources
- reducing inequities and inequalities
- screening
- assessing the impact of interventions

The paper concludes with an infographic that summarizes the validity of the *EYE*.

III. Valid Assessments of Early Child Development

As educators make the shift towards a systematic phonics-based approach to teaching, they are questioning whether the traditional reading assessments meet their needs. Many reading assessments are outdated and inconsistent with the rapidly changing research on the science of reading. This section describes nine criteria for a valid assessment and discusses the validity claims, research, and validity arguments for the *EYE*.

Policy Coherence

> Policy Coherence is "the systematic promotion of mutually reinforcing policy actions across government departments and agencies creating synergies towards achieving the agreed objectives." [10]

Policy coherence is arguably the most important criterion of a reading assessment as it requires educators to consider whether the assessment is coherent with the jurisdiction's policies and long-term objectives.

Validity Claim. The *EYE* is aligned with provincial policies and curricula on early childhood education.

Research. Educational policy entails setting goals and developing a course of action for achieving them. The 'course of action' requires identifying a set of strategies for achieving the outcomes and a plan for their execution. The *EYE* is based on a policy framework, *Educational Prosperity*, as described in the UNESCO report *Learning Divides*.[11] The report emphasizes three questions relevant to policy coherence:

- What strategies are most likely to lead to system improvement?
- How can resources be allocated to strengthen the foundations for students' success and develop a more equitable school system?
- How can monitoring data be used to assess the effects of policies that change one or more of the structural features of the school system?

---

[10] Morales, E. S., & Lindberg, C. (2017, July). *Tracking progress on policy coherence for sustainable development at the national level: What and how to measure?* Coherence for sustainable development. https://www.oecd.org/gov/pcsd/Coherence%20for%20Development_Issue_9.pdf

[11] Willms, J. D. (2018). *Learning Divides: Using Monitoring Data to Inform Education Policy*. Montreal: UNESCO Institute for Statistics.

Validity Argument. The *EYE* is being used to help provinces and school divisions transition from a 'three-cueing' approach to teaching reading to a systematic, phonics-based approach. The Learning Bar's course, *Systematic Reading Instruction*, which accompanies the *EYE*, describes prerequisite reading skills and a three-phase model for teaching reading during the early and late primary school years.

Feasibility

> The feasibility of an assessment is concerned with whether it can be successfully implemented in a jurisdiction based on technical, legal, financial, political, operational, and contextual considerations.

Technical feasibility considers whether the necessary technology, infrastructure, and support are available for its implementation.

Legal feasibility considers whether the implementation of an assessment aligns with a jurisdiction's laws and regulations.

Financial feasibility asks whether an assessment is cost-effective and financially sustainable, given a jurisdiction's budget and funding sources.

Political feasibility assesses whether an assessment has the support of key stakeholders, including teachers, district and Ministry administrators, parents and caregivers, and community leaders.

Operational feasibility examines the human resources, and organizational capacity that are necessary for successful implantation.

Contextual feasibility considers the broader socio-economic and cultural context in which the assessment will be implemented.

Validity Claim. The *EYE* can be successfully implemented in North American and Australian provinces and states, meeting their technical, legal, financial, political, operational, and contextual requirements.

Research. The Learning Bar uses the framework described by Donghun Yoon, which includes the key elements described above.[12]

A perennial issue among early childhood educators pertains to competing views about how children learn during the early years. One view, rooted in constructivism, stresses the importance of active play.[13, 14, 15] The teacher's role is to create an environment where children can learn and make discoveries through active play while helping them construct knowledge through guiding interactions and dialogue. A competing perspective is that children need to acquire specific skills

---

[12] Yoon, D. (2021). Policy feasibility analysis model. In *Preliminary Feasibility for Public Research and Development Projects* (pp. 49-57). Emerald Publishing Limited.

[13] Bronfenbrenner, U. (1977). Toward an experimental psychology of human development. *American Psychologist, 32,* 513-532.

[14] Piaget, J. (1952). *The Origins of Intelligence in Children.* New York: International University Press.

[15] Vygotsky, L. S. (1978). *Mind in Society.* Cambridge, MA: Harvard University Press.

that will enable them to succeed. Many children with disabilities and those who grow up in impoverished environments do not have the same opportunities for active play and social interactions as their peers from higher socio-economic backgrounds. They require learning activities that are specifically designed to build foundational literacy and numeracy skills.[16]

The Learning Bar's position concerning the *EYE* is that the two competing perspectives set up unhelpful false dichotomies such as 'child-centered' versus 'teacher-centered' or 'play-based' versus 'work-based.' Active play is essential in preschool and kindergarten programs, but its value does not diminish when children enter Grade 1 or Year 1. Similarly, even very young children can benefit from structured lessons. They not only provide an efficient way to learn pre-literacy skills, but they also support learning rules and routines and increasing self-regulation.[17, 18]

Validity Argument. The *EYE* is being used province-wide in Saskatchewan, Manitoba, and New Brunswick, territory-wide in Yukon Territory, and in 22 school divisions in Alberta. It has also met the feasibility requirements in several other countries.

The *EYE* is embraced by most teachers for several reasons. One is that the domains of the *EYE-DA* and *EYE-TA* are explicitly linked to structured lessons that include active play-based learning activities. The pre-reading skills assessed with the *EYE-PR* are linked to activities pertaining to concepts about print, letter knowledge, sound-letter relationships, and phonological and phonemic awareness.

Some teachers feel that assessments displace valuable teaching time, restrict what is taught and learned, and are used to hold teachers accountable in unfair ways. These concerns are largely mitigated by the alignment of the *EYE* with provincial and state curricula, the usefulness of the data for informing classroom instruction, and the immediate availability of the data. The *EYE* data are not used to make comparisons among schools or school districts. Instead, the emphasis is on determining how each child is doing, what they need to learn next, and how well they are progressing.

---

[16] Linan-Thompson, S. (2014). *Screening of school readiness skills: A review of the literature*. Washington: US AID.

[17] Adamson, D. R. (2010). Classroom management: 24 strategies every teacher needs to know. New York: Scholastic.

[18] McClelland, M. M., & Tominey, S. L. (2016). *Stop, think, act: Integrating self-regulation in the early childhood classroom*. New York: Routledge.

Sustainability

Sustainability pertains to whether an assessment can be conducted consistently for a prolonged period with sustained benefits.

An assessment is sustainable when its use is embedded in the culture of the school and consistently provides reliable, accurate, and beneficial results.

Validity Claim. The use of *EYE* is embedded in the culture of schools and school divisions.

Research. An assessment is sustainable when teachers have an opportunity to continually refine their assessment practices, its administration is practical in the school context, and the test publisher is responsive to teacher feedback.[19, 20]

Validity Argument. Provincial contracts for the *EYE* have been regularly renewed, in many cases for longer than ten years. The Learning Bar conducts an annual survey of its users, which is followed by a review of its items and administration. One of the most significant changes stemmed from teachers' requests for a shorter version of the *EYE-TA*. In 2022, The Learning Bar developed and released a shorter 24-item version of the assessment, which did not unduly affect its reliability or the ability to monitor long-term trends.

[19] Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in Practice: Developing Language Assessments and Justifying their Use in the Real World*. Oxford University Press.

[20] Shepard, L. A. (2009). The role of assessment in a learning culture. Journal of Education, 189(1-2), 95-106.

Educational Effect

> The educational effect of an assessment refers to its impact on student learning that occurs through changes in teachers', students', and parents' and caregivers' behaviours.

Assessment drives student learning. It can shift the emphasis teachers place on particular learning outcomes, the design of their daily lessons, and the time devoted to certain topics. When an assessment is aligned with the curriculum, the adage that 'teachers teach to the test' becomes 'teachers teach to the curriculum.' The educational effect is maximized when an assessment can be easily administered at regular intervals throughout the school year and quickly provides usable reports.

Validity Claim. The use of the *EYE* significantly improves student learning outcomes by positively influencing the behaviors of teachers, students, parents, and caregivers.

Research. Activity theory posits that people mediate their behaviour in response to various tools and signs.[21] Assessments influence student outcomes indirectly through changes in teaching practices, the design of lessons, and learning materials.

A study of the *EYE-TA* results for children attending the CHANCES Smart Start Program on Prince Edward Island (PEI) found that the pre-post gains in developmental outcomes increased significantly over a five-year period.

The Learning Bar conducted a national evaluation of the effects of kindergarten programs in Uruguay. The sample comprised 22,582 children aged 4 who attended ECEC programs in 1,591 classrooms in Uruguay's 19 departments (regions). [22] The researchers found that the levels of annual growth in cognitive and language skills varied substantially within and among classes but were positive in most schools. Moreover, annual growth exceeded maturity effects; that is, it exceeded the effects one would expect if the children had not attended a kindergarten program.

Validity Argument. The *EYE* has a measurable and sustained benefit on children's early years outcomes. The sustained improvement in results is due mainly to teachers' comprehension of the *EYE* framework and their application of the associated structured lessons. Findings from the CHANCES study and those of several school divisions suggest that the positive effects of the *EYE* are realized after the first year.

---

[21] Gibbs G., & Simpson C. *How Assessment Influences Student Learning – a Conceptual Overview.* Heerlen, the Netherlands: Open University. Centre for Higher Education Practice 2002.

[22] Lopez, A. Y., & Willms, J. D. (2020). A national evaluation of kindergarten outcomes: Findings from Uruguay. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International Perspectives in Educational Effectiveness Research* (pp. 159-176). Springer.

Equivalence and the Absence of Bias

> Equivalence asks whether test scores obtained for different sub-populations can be interpreted in the same way. When test scores are not equivalent across groups, comparisons of test scores can be biased.[23]

Educational leaders at all levels recognize that social and economic development requires the universal provision of education from early childhood to adolescence. Consequently, they set policies and develop interventions to increase all students' outcomes while reducing inequalities among sub-populations. We use the term 'sub-populations' to refer to students in various demographic groups, including gender, disability, social class, ethnicity, national origin, immigration status, sexual orientation, and religion.[24]

A key use of a reading assessment is to estimate the magnitude of differences among subpopulations and monitor gaps in achievement over time. However, when test scores are not equivalent among subpopulations, comparisons can be biased. [25]

Validity Claim. The *EYE* test items can be assessed to eliminate cultural bias, ensuring fair and valid measurement of preschoolers' abilities across diverse cultural groups.

Research. All types of assessments, from the formative assessments teachers use in a lesson to the formal tests used in provincial or state examinations, comprise a set of assessment tasks. An assessment task is a request to do something that will be evaluated. We consider three elements of an assessment task:

- a request – what a student is asked to do
- the performance – how a student responds
- the evaluation – a judgment about the quality of the performance

Bias can be introduced when any of these elements differs among subpopulations. Standardizing administration procedures reduces bias. For example, with the *EYE-PR*, students hear the request from the same person using an online application. The classroom teacher's evaluation is also standardized with a clearly defined scoring procedure.

---

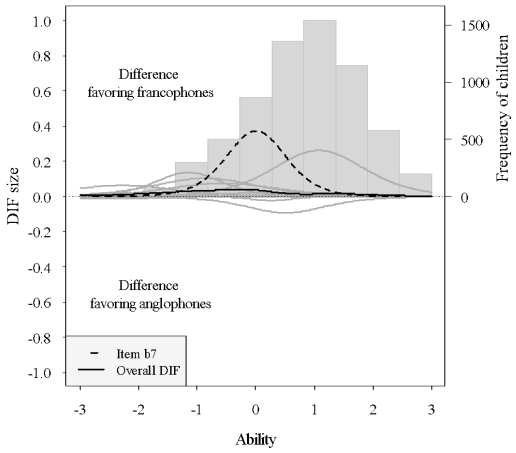
[23] van de Vijver, F. J. R., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology = Revue Européenne de psychologie appliquée,* 47(4), 263-280.

[24] Willms, J. D. (2009a, October). "Classroom diversity and Inclusion: The Educational Advantage"**.** Plenary presentation at the Return to Salamanca – Global Conference on Inclusive Education. Salamanca, Spain.

[25] He, W., & Wolfe, E. W. (2010). Item equivalence in English and Chinese translation of a cognitive development test for preschoolers. *International Journal of Testing*, *10*(1), 80-94.

The Learning Bar investigated item bias in the *EYE-DA* using a statistical technique called Differential Item Functioning (DIF).[26] The research identified a small number of items that exhibited bias, meaning that children from different cultural backgrounds did not have an equal chance of answering these items correctly, even if they had the same underlying ability level. In response, the items were revised to reduce bias.

This figure shows the Differential Item Functioning (DIF) for cognitive skills assessed with the *EYE-DA*. The histogram shows the distribution of cognitive skills for the domain. Item B7, which assesses a student's ability to count 15 identical objects, favors Francophone students at the lower end of the ability scale. This means that Francophone students are more likely to display mastery of this skill than their Anglophone counterparts, given their overall ability in this domain. The solid black line shows the total DIF for the domain. While it favours Francophone students, it is very small.



Validity Argument. The Learning Bar has developed a rigorous approach to assessing bias in its test items. In many provinces and school divisions, the interest is in estimating the gaps in achievement between males and females, Anglophone and Francophone students, and Indigenous and non-Indigenous students. This requires the collection of demographic data alongside the *EYE* data, which is not done routinely in all jurisdictions.

The items of the *EYE-PR* will be examined using this approach in the fall of 2024.

[26] Marotta, L., Tramonte, L., & Willms, J. D. (2015). Equivalence of testing instruments in Canada: Studying item bias in a cross-cultural assessment for preschoolers. *Canadian Journal of Education, 38*(3), 1-23.

Reliability

> Reliability refers to the consistency of a measurement process across varying testing occasions, evaluators, and contexts.[27]

Validity Claim. The *EYE* yields reliable scores of students' early developmental and pre-reading skills. The items of the *EYE* have an appropriate level of difficulty and can reliably identify children who are vulnerable.

Research. Reliability refers to the consistency of a measurement process. Does an instrument yield similar results if it is administered by different assessors, on different occasions, or in different settings? Reliability is closely related to validity because acquiring evidence about the consistency of measurement requires that the observed test items are valid indicators of the underlying construct.

The reliability of assessments depends on several factors. Tests are more reliable when:

- they include a sufficient number of test items
- administration procedures are standardized
- scoring procedures are exacting and clear
- testing conditions are consistent
- the test items are objective and of high quality

Reliability coefficients in the range of 0.70 to 0.80 are usually considered acceptable for most purposes.[28] Test developers must ensure that a test is sufficiently reliable for its intended purpose and yet short enough to be practical for classroom use.

Validity Argument. The *EYE* has stringent guidelines for its administration and scoring, including, for example, how items are presented, the time allotted for a student to respond, and whether a prompt can be given when a student hesitates to respond. When most teachers are introduced to the *EYE*, they receive a comprehensive training session from The Learning Bar.

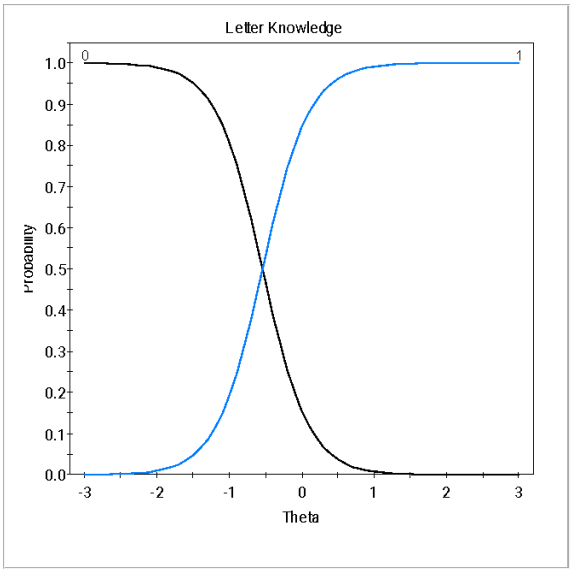
The Learning Bar uses a technique called Item Response Theory (IRT) to ensure the quality of every test item. The technique enables the researcher to estimate the relative 'difficulty' of a test

[27] Livingston, S. A. (2018). *Test reliability—Basic concepts* (Research Memorandum No. RM-18-01). Princeton, NJ: Educational Testing Service.

[28] Standards for Educational and Psychological Testing (2014). American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME).

item and how well each item 'discriminates' between children who have weak versus strong skills. It also ensures that the assessment captures a wide range of abilities.[29]

The figure shows the item response curve for the *EYE-PR* item on letter knowledge, based on a Rasch 2-category model. In this example, the graph indicates the probability that a student would achieve mastery – a score of 3 or 4 – on the 4-point rubric. The results indicate that, on average, about 80% of students achieved mastery of this task. The IRT analysis also revealed that about 38% of students exceeded expectations on this task as they could correctly identify at least 24 of the 26 lowercase letters.



The Learning Bar estimates the reliability of the *EYE* based on its Canadian database. The reliability estimates, estimated by Cronbach's alpha, have been remarkably stable for over a decade. The number of test items and the reliability coefficients are shown in the table below.

[29] Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14*, 97-116.

Reliability of the EYE

	EYE-DA		*EYE-TA*		*EYE-PR*	
	No. Items	Reliability	No. Items	Reliability	No. Items	Reliability
Awareness of Self and Environment	12	0.89	8	0.91		
Social Skills		-	8	0.94		
Cognitive Skills	12	0.86	8	0.89		
Language	12	0.91	8	0.91		
Physical Development	13	0.88	10	0.86		
Pre-reading skills					10	0.87
Full Scale	49	**0.96**	42	**0.96**	10	**0.87**

The *EYE* provides reliable results as a full evaluation system, which includes three separate assessments. These assessments provide reliable results for a wide range of skills, for students from pre-kindergarten to Grade 1 or Year 1.

Content Validity

> Content Validity is the extent to which an assessment covers a representative sample of a curriculum's learning objectives.

The design of traditional assessments is based on a content framework that sets out the learning objectives and the domains of content. The test developers then select a sample of learning objectives from each domain and prepare test items to measure the skills of the sampled objectives.

Validity Claim. The *EYE* covers a representative sample of the learning objectives of most provincial and state curricula for preschool, kindergarten, and Grade 1 or Year 1.

Research. The UNESCO *Education for All Global Monitoring Report* noted that the consensus from research is that:

> School readiness encompasses development in five distinct but interconnected domains: language development, cognitive development, social and emotional development, approaches to learning, and physical development.[30]

A position statement of the National Association for the Education of Young Children supported this view:

> Chief among the professional responsibilities of early childhood educators is the responsibility to plan and implement intentional, developmentally appropriate learning experiences that promote the social and emotional development, physical development and health, cognitive development, and general learning competencies of each child served.[31]

The pre-reading skills of most curricula emphasize the constructs recommended by the U.S. National Reading Panel: phonics, (phonological awareness and letter-sound relationships), phonemic awareness, fluency, vocabulary, and comprehension. [32]

Validity Argument. The selection of domains for the *EYE-DA* and the *EYE-TA* was influenced by the goals of the U.S. National Education Goals Panel and by a report on national school readiness

---

[30] UNESCO (2006). *Strong foundations: Early childhood care and education.* Paris: UNESCO.

[31] National Association for the Education of Young Children (NAEYC). (2012). *Position Statement: Principles of Child Development and Learning and Implications That Inform Practice*. NAEYC.

[32] National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). National Institute of Child Health & Development.

indicators developed by a 17-state partnership.[33, 34] The selection strived to meet three conditions. First and foremost, they had to be useful for teachers in their daily practice: for creating learning activities, planning daily activities, and providing a systematic framework to structure observations and conduct informal assessments. Second, they needed to be consistent with differing approaches to learning and philosophies of pedagogy. Finally, they had to be consistent with the current literature on early childhood development.

The content of the *EYE-PR* covers the core pre-reading skills of the National Reading Panel.

[33] Shepard, L., Kagan, S. L., & Wurtz, E. (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: National Education Goals Panel.

[34] Rhode Island Kids Count (Organization), (2005). *Getting ready: Findings from the National School Readiness Indicators Initiative: A 17-state partnership*. Providence, RI: Rhode Island KIDS Count. Retrieved from http://www.aecf.org/m//resourcedoc/RIKC-GettingReady-2005.pdf.

Criterion-related Validity

Test scores are correlated with scores based on previously validated tests and with competencies or events in the future. Criterion-related validity is often separated into concurrent and predictive validity.

Evidence supporting claims of concurrent validity is derived from two or more assessments taken at the same time. For example, the validity of a test purporting to assess children's receptive vocabulary skills would be strengthened by examining whether its scores correlated with a more extensive test of receptive vocabulary, such as the *Peabody Picture Vocabulary Test* and the *Expressive Vocabulary Test.*

Evidence supporting claims of predictive validity is usually derived from longitudinal studies that follow a cohort of students. For example, Scarborough's seminal research summarized the results of 61 longitudinal studies that followed cohorts of students from early primary school to their first or second grade.[35] She found that children's letter identification and phonological awareness skills when they started early primary school were predictive of their ability to decode one or two years later. These skills are also predictive of whether students become competent readers at the end of primary school and at age 15.[36]

Validity Claim. The *EYE* has high criterion-related validity, evidenced by strong correlations with established literacy tests and future literacy outcomes.

Research. A longitudinal study of the predictive validity of the *EYE-TA* was conducted in five New Brunswick school districts. Over 11,000 kindergarten children were assessed on the five measures of the *EYE-TA.* At the end of Grade 2, these children completed the provincial assessment tests of oral and written reading ability. The *EYE-TA* proved to be a strong predictor of children's later reading skills, especially their skills in cognitive and language development.

A study of the concurrent validity of the *EYE-PR* was conducted in 2024. Ten of the items of the *EYE-TA* measures are predictive of students' later reading skills. These include five cognitive items and five language and communication items:

This child can:

---

[35] Scarborough, H. S. (1989). Prediction of reading disability from familial and individual differences. *Journal of Educational Psychology*, *81*(1), 101-108.

[36] Adlof, S. M., Catts, H. W., & Lee, J. (2010). Kindergarten predictors of second versus 8th grade reading comprehension impairments. *Journal of Learning Disabilities, 43*(4), 332-345.

- recognize and name 15 letters of the alphabet
- understand how books are manipulated
- match each letter (A, D, G, M) with the picture beginning with that letter
- tell the first letter and letter sound in their name and in ‘fish,’ ‘king,’ and ‘sun.’
- recognize and name the numerals 1 to 10
- follow two-step commands (e.g., "go to your table and open the book")
- listen to and comprehend classroom stories read by the teacher
- understand the meaning of action words (e.g., hop, spin, march, clap, etc.)
- use picture cues to tell a story in proper sequence with minimal prompting
- predict what will happen next in a story

The mean scores on the *EYE-PR* were significantly correlated ($r = 0.66$, $p<.001$) with the mean scores for the ten *EYE-TA* items, and with each of the *EYE-TA* items. The strongest correlations were with the *EYE-TA* items on letter knowledge ($r=0.504$, $p<.001$), matching ($r=0.610$, $p<.001$), and identifying initial sounds ($r=0.576$, $p<.001$).

Validity Argument. The *EYE-TA* is strongly correlated with reading scores at the end of Grade 2, and the *EYE-PR* is strongly correlated with teachers’ ratings of the ten items most strongly related to pre-reading skills. Taken together, these items support the validity claim that the *EYE* has high criterion-related validity. The findings are consistent with the long-standing findings of Scarborough.

IV. Use Validity of the Early Years Evaluation

The *EYE-TA* has multiple uses which are listed in Table 1, grouped into five categories. The validity claims, relevant research, and validity arguments follow the table.

Uses of the *EYE*

I. Improve Classroom Instruction

- Guide planning of classroom activities to develop children's pre-literacy skills
- Increase teachers' repertoire of high-yield teaching strategies
- Strengthen partnerships with families and community agencies

II. Assess the Effectiveness of ECEC and Kindergarten Programs

- Predict levels of literacy skills at age 8 or 9
- Assess pre-post learning gains in pre-literacy skills
- Assess the effects of school or jurisdiction-wide interventions

III. Reduce Inequalities of Sub-populations

- Estimate inequalities in pre-literacy skills upon school entry
- Estimate the extent of reduction in inequalities in pre-literacy skills during kindergarten

IV. Screening

- Identify students requiring extra support
- Estimate the prevalence of students with diverse learning needs

V. Allocate Educational Resources

- Assess progress towards meeting strategic goals
- Allocate resources in an efficient manner

Improving Classroom Instruction

Validity Claim. The *EYE* improves classroom instruction by enabling teachers to plan effective classroom activities, increase their repertoire of high-yield teaching strategies, and strengthen partnerships with families and community agencies.

Research. The primary role of assessment in early childhood education is to gather comprehensive, accurate information about children's development that informs and enhances teaching practices.[37] The "intentional teacher" uses assessment to identify areas in which children can benefit from particular learning opportunities. [38]

Kindergarten programs that explicitly teach key pre-literacy skills can enable children with differing ability and family backgrounds to become successful readers.[39] The term 'high-yield teaching strategies' refers to strategies supported by research that significantly improve assessed student achievement.[40] They are linked to long-term goals and consistent with theories about how children learn.

Early childhood assessments can play a pivotal role in strengthening partnerships with families and community leaders by promoting collaboration and establishing shared goals.[41] The overarching goal is to connect teachers with parents and caregivers so they are partners in their children's development. [42] Early childhood assessments can achieve that by regularly sharing results with families to keep them informed about their child's progress, providing clear, understandable reports of children's strengths and areas for improvement, and setting learning

[37] Snow, C. E., & Van Hemel, S. B. (2008). *Early Childhood Assessment: Why, What, and How*. National Research Council.

[38] Epstein, A. S. (2007). *The Intentional Teacher: Choosing the Best Strategies for Young Children's Learning*. National Association for the Education of Young Children (NAEYC).

[39] Schatschneider, C.F., Fletcher, J., Francis, D., Carlson, C., & Foorman, B. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96*(2), 265-282.

[40] Lezotte, L. W. (1997*). Learning for All*. Okemos, MI: Effective Schools Products Ltd.

[41] Henderson, A. T., Mapp, K. L., Johnson, V. R., & Davies, D. (2007). *Beyond the Bake Sale: The Essential Guide to Family-School Partnerships*. The New Press.

[42] Epstein, J. L., & Sanders, M. G. (2006). *Connecting Home, School, and Community: New Directions for Social Research*. Handbook of the Sociology of Education.

goals based on assessment data fosters a sense of shared responsibility and partnership.[43, 44,]

Validity argument. The *EYE* plays a key role in improving classroom instruction. It provides a systematic framework for teachers to structure their observations and conduct formative assessments. It also allows them to structure their child-centered environments to ensure that each child's specific learning needs are met. Teachers using the *EYE* maintain a focus on key developmental outcomes.

Two unique aspects of the *EYE* are its explicit links to professional development and classroom learning activities that use high-yield teaching strategies.

The Learning Bar's course on Systematic Reading Instruction is designed to provide teachers with an understanding of systematic reading instruction and the research underlying it and the tools to apply the key concepts and strategies of systematic reading instruction in their classrooms.

Two sets of classroom lessons with structured lessons accompany the *EYE*. The *EYE-100* includes forty activities that emphasize skills in the cognitive domain, forty activities that emphasize skills in the language and communication domain, and twenty activities aimed at strengthening students' inquiry problem-solving skills. The *EYE-PR* is linked to a set of activities designed to teach the core pre-reading skills: concepts about print, letter knowledge, phonological awareness, phonemic awareness, and sound-letter relationships.

Child reports. Individual child reports are often used in the first parent-teacher meetings to discuss a child's strengths and areas where further experience would benefit them.

The child reports are provided in 21 different languages, including six Indigenous languages.

Classroom reports. The classroom reports provide detailed information on the RTI results for each child, and a summary for the students in the class.

Sample reports are provided in Appendices A and B.

Many school districts share the summary reports of *EYE* results and maps to help community agencies identify areas of need. This practice strengthens partnerships across government agencies and between schools and local organizations.

---

[43] Weiss, H. B., Lopez, M. E., & Rosenberg, H. (2010). *Beyond Random Acts: Family, School, and Community Engagement as an Integral Part of Education Reform*. Harvard Family Research Project.

[44] Henderson, A. T., Mapp, K. L., Johnson, V. R., & Davies, D. (2007). *Beyond the Bake Sale: The Essential Guide to Family-School Partnerships*. The New Press.

To summarize, the validity of the claim that the *EYE* improves classroom instruction is supported by its provision of detailed and actionable reports that enable teachers to plan targeted and effective classroom activities, expand their use of high-yield teaching strategies, and foster meaningful partnerships with families and community agencies. The *EYE* is a comprehensive approach to assessment that improves classroom instruction, leading to improved outcomes for young children.

Assessing the Impact of Interventions

Validity Claim. The *EYE* can be used to assess the effectiveness of preschool and kindergarten programs.

Research. Two broad conclusions have emanated from research on the effectiveness of preschool and kindergarten programs: attending high-quality programs leads to better schooling outcomes and program effects tend to be stronger for children from disadvantaged backgrounds and those with lower ability.[45, 46]

A meta-analysis examining the effects of 16 school readiness programs for preschool children found an effect size of 0.93.[47] Research has also found that early childhood programs designed by researchers had average effect sizes for cognitive development of 0.57; while state and local public-school programs had average effect sizes of 0.32.[48]

Validity Argument. When assessing the effectiveness of preschool and kindergarten programs with the *EYE-TA*, jurisdictions typically examine whether there are increases, from the beginning to the end of the school year, in the percentage of children that have 'appropriate development rather than having 'moderate' or 'significant difficulties.'

Our approach to quantifying the effects of interventions is to use a 'months of schooling' metric. We have accurate estimates of the 'maturity effect' for students' growth in each of the five *EYE-TA* domains; that is, we know the average growth rate for students if they are not attending school. We also have estimates for the average growth rate in each domain for a large representative sample of Canadian students and schools. Thus, for a district implementing an intervention, we can indicate whether fall-to-spring growth increased because of the intervention. However, The Learning Bar's emphasis has been on examining year-over-year changes in effects, rather than making comparisons to a national standard. The Learning Bar has also been able to estimate the effects on performance for students who have had prolonged absences from school.

---

[45] Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., Ludwig, J., Magnuson, K., Phillips, D., & Zaslow, M. (2013). *Investing in Our Future: The Evidence Base on Preschool Education.* Washington, DC: Society for Research in Children and Foundation for Child Development.

[46] Barnett, W.S. (2011). Effectiveness of early educational intervention. *Science*, *333*, 975-78.

[47] Çoban, A., Yildiz, E., Aslan, H. S., Ögütcen, A., & Koca, Ö. (2023). Effectiveness of school readiness programs for preschool children: A meta-analysis study. *Educational Policy Analysis and Strategic Research*, *18*(3), 50-77.

[48] Kay, N., & Pennucci, A. (2014). *Early Childhood Education for Low-Income Students: A Review of the Evidence and Benefit–Cost Analysis (Doc. No. 14-01-2201).* Olympia: Washington State Institute for Public Policy.

Reducing Inequalities and Inequities

Validity Claim. The *EYE* can be used to assess the extent of inequalities between males and females, between Indigenous and non-Indigenous students, and students from varying socioeconomic backgrounds.

Research Background. The Learning Bar distinguishes between 'equality' and 'equity' following the definitions used by the OECD and the Inter-American Development Bank.[49, 50]

- 'equality' refers to differences among sub-populations in the distribution of their educational outcomes
- 'equity' refers to differences among sub-populations in their access to school and to the resources and schooling processes that affect schooling outcomes.

For example, a school district may use *EYE-PR* data to determine the gaps in pre-reading skills between males and females when they enter Grade 1 or Year 1. That is equality. It may also assess how educational resources are inclusively allocated among schools. That is equity.

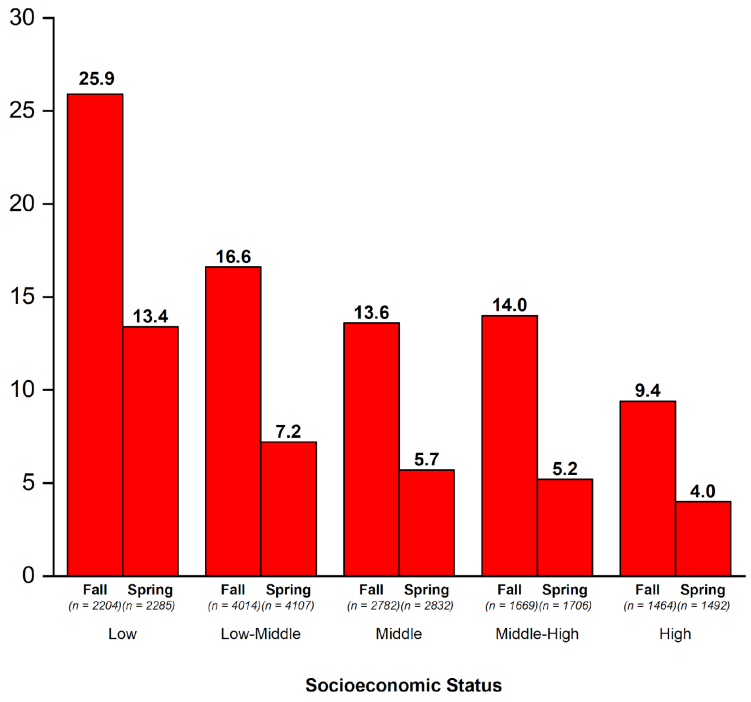
This distinction between equality and equity is important because measures of equality indicate the magnitude of differences among subpopulations – how big is the problem? – while measures of equity call for policies that address the problem – what needs to change to reduce inequalities?

Validity Argument. The *EYE* includes the collection of demographic data for children's sex and their postal codes. Some jurisdictions also collect data on Indigenous status. The postal code data are linked to Canadian Census data at the Enumeration Area level. The Learning Bar used the data to develop an array of demographic characteristics, including an overall measure of socioeconomic status (SES). The figure below shows the distribution of students with Tier 3 learning needs for the full Canadian population. The results show the reduction in vulnerability from the beginning to the end of the school year.

---

[49] Willms, J. D. (2011). *An analysis plan on educational equality and equity: Recommendations for the OECD Education at a Glance.* Paper prepared for the OECD NESLI INES Network for the collection and adjudication of system-level descriptive information on educational structures, policies and practices (NESLI). The Learning Bar.

[50] Willms, J. D., Tramonte, L., Duarte, J. & Bos, S. (2012). Assessing educational equality and equity with large-scale assessment data: Brazil as a case study. Washington: *Inter-American Development Bank.*



Percentage of Canadian children with Tier 3 learning needs by SES quintile, 2017-18

Screening

Validity Claim. The *EYE* accurately identifies children at risk for learning difficulties, enabling timely and targeted interventions at the school and district levels.

Research Background. Screening assessments provide data that can be used to identify children at risk for reading difficulties. Early detection allows for timely intervention that can prevent long-term literacy problems. With early identification, educators can implement preventive measures that mitigate the development of severe reading difficulties. This proactive approach is more effective and less costly than remedial education.[51]

A screening tool can also be used to identify students who require further assessments to identify specific developmental problems that may affect their learning.

Validity Argument. The *EYE* accurately identifies children at risk for learning difficulties, enabling timely and targeted interventions.

Immediately upon completing the *EYE-TA*, teachers receive a report based on a Responsive-Tiered Instruction (RTI) framework. RTI calls for a tiered approach to instruction consisting of continuous assessment of children's progress in the regular classroom setting and additional support for children who have learning difficulties or behavioural challenges.[52] The Learning Bar uses the term Responsive-Tiered-Instruction because it uses data from the *EYE* as *leading* indicators, enabling teachers to identify children who are likely to need Tier 2 or Tier 3 instruction before they enter Grade 1. The three-tier RTI classification data is based on a prediction model that takes into account the age of the child at the time of assessment and their skill level in the five domains. The model, which is based on longitudinal data, weights some skills more heavily than others. Skills in the cognitive and language domains figure most prominently, but children's

[51] Schatschneider, C., Petscher, Y., & Williams, K. M. (2008). How to evaluate a screening process: The vocabulary of screening and what educators need to know. In L. M. Justice & C. Vukelich (Eds.), *Solving problems in the teaching of literacy. Achieving excellence in preschool literacy instruction* (pp. 304-316). New York, NY, US: Guilford Press.

[52] Gersten, R., Compton, D., Connor, C.M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W.D. (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide.* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

awareness of self and environment, their social skills and approaches to learning, and their fine motor skills also contribute to the classification.

In most school districts, all students are assessed at the beginning of their kindergarten year, and those classified as having Tier 2 or Tier 3 learning needs are assessed again near the end of the school year, before they enter Grade 1 or Year 1. Kindergarten teachers use the RTI classification in combination with the domain-specific results to plan their instruction. Grade 1 teachers use the results to plan for the resources they will need for the upcoming year and ensure that children who score Tier 3 have the necessary further assessments for developing an individualized learning plan.

The Learning Bar advocates using the *EYE-PR* in combination with the *EYE-TA*. The *EYE-TA* was designed as a screening assessment to identify vulnerable children, and as such it does not provide information across the full range of abilities. The *EYE-PR* covers pre-reading skills and provides reliable information across the full range of abilities. Teachers can use the *EYE-PR* to provide additional information on students with Tier 1 learning needs.

The combination is ideal for Ministry-level initiatives, as it provides reliable information on vulnerability levels and accurate data for tracking changes in the full student population.

Allocating Resources

Validity Claim. The *EYE* data can be used to inform decisions about allocating material and human resources.

Research Background. A key role of district and school administrators concerns the allocation of funds for staffing, professional learning, instructional material and technology, and support for students with diverse learning needs. [53]

Some examples include:

Staffing

- Hiring teachers, support staff, and specialists based on enrollment and student needs.
- Assigning staff to specific roles, classrooms, and programs to optimize the use of their skills and expertise.

Inclusive Education Needs

- Providing additional support and resources to students who are at risk or have diverse learning needs.
- Supporting intervention programs for students who need additional academic or behavioural support.

Professional Learning

- Providing professional development opportunities to improve staff skills and effectiveness.

Material Resources

- Allocating funds for instructional materials, technology, and classroom supplies.
- Implementing programs and initiatives that promote diversity, equity, and inclusion.

Evaluating students' skills when they enter school is central to developing a set of strategies and making plans for their execution. It also requires policies about how best to allocate available resources.

[53] Kowalski, T. J. (2013). *The School Superintendent: Theory, Practice, and Cases*. SAGE Publications.

Validity Argument. The *EYE* is used to inform decisions about resource allocation by accurately portraying how children with diverse learning needs are distributed within and between schools. The data can support a demand-side approach to diverse learning needs funding that overcomes the problems associated with flat-grant and supply-side funding models.

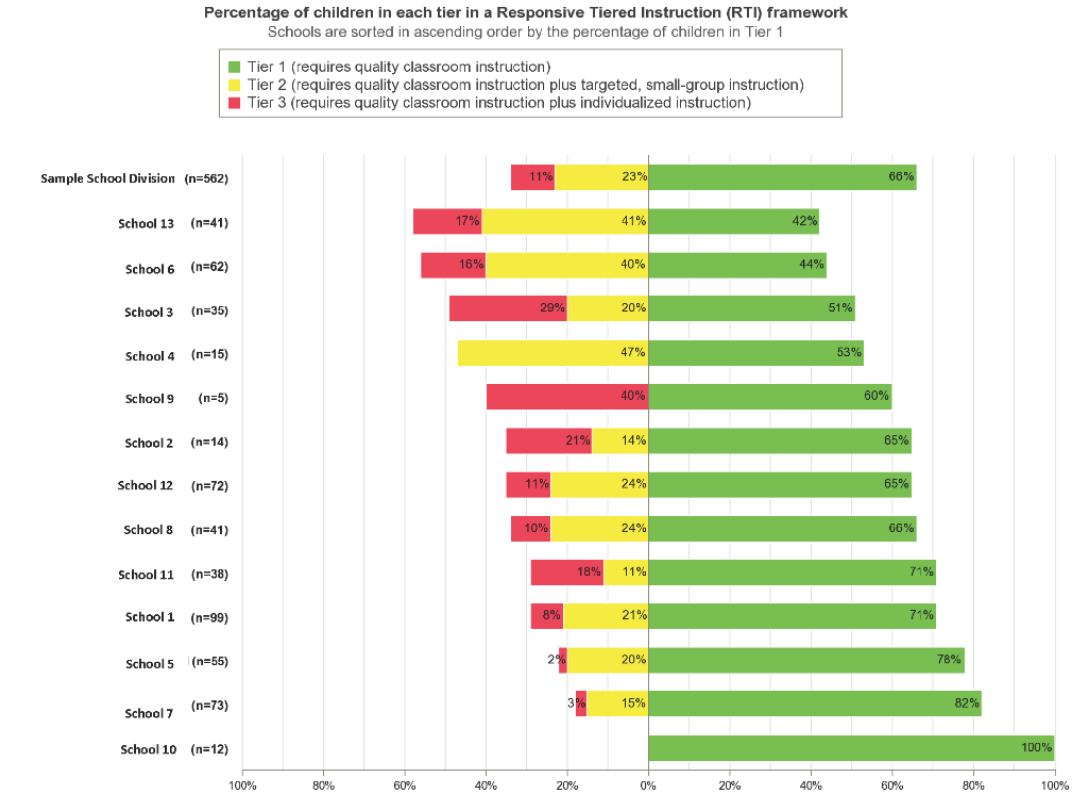
The 50-20 Rule**.** Analyses of the *EYE* data have consistently found that in most school districts, 50% of the students with diverse learning needs are in 20% of the district's schools. For example, in a school district with 25 schools, about one-half of its students with Tier 2 and 3 learning needs are in 5 of its schools. These findings have implications for how resources for teachers, educational assistants, support staff, and specialists are allocated to schools.

They also have implications for strategies or interventions to improve the performance of students with diverse learning needs. Willms described five types of strategies that can be implemented in a school district to improve student outcomes.[54, 55] Two strategies are especially relevant to the use of *EYE* data: universal and performance-targeted strategies. A universal strategy strives to improve the outcomes of all students in a jurisdiction, while a performance-targeted strategy expends more resources on programs and services for students with diverse learning needs.

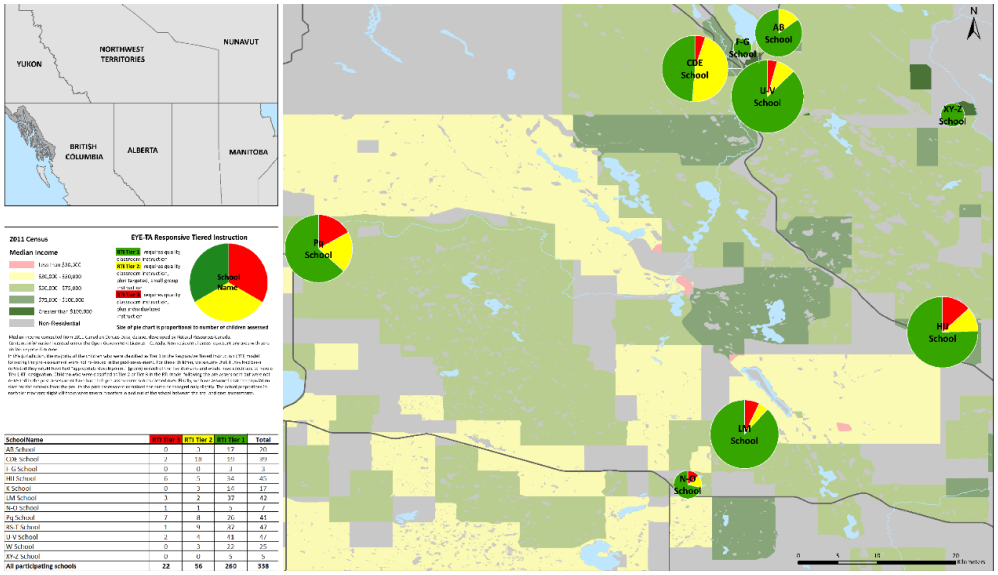
The summary reports and the maps of the *EYE* data allow school administrators to identify the schools with a high percentage of students with diverse learning needs. Two examples are shown below.

[54] Willms, J. D. (2010). School Composition and Contextual Effects on Student Outcomes. *Teachers College Record, 112*(4), 1008-1037.

[55] Willms, J. D. (2018). *Learning Divides: Using Monitoring Data to Inform Education Policy*. Montreal: UNESCO Institute for Statistics.



Prevalence of children in in each tier of a Responsive Tiered Instruction Framework (sample school district)



Map of the median income and prevalence of children in each tier of a Responsive Tiered Instruction. (sample school district)

Inclusive Education Funding. School districts and governments are inconsistent in their approaches to the funding and delivery of educational services to students with diverse learning needs. Many jurisdictions use a combination of approaches for funding diverse learning education programs and services. These include elements of the following approaches:

- flat grant models allocate funds on a per capita basis using the total student enrolment of a school jurisdiction;
- supply-side models provide reimbursement for the amount a school jurisdiction spends on inclusive education services; and
- demand-side models allocate funding based on the number of children in a school jurisdiction deemed to have diverse learning needs.

The Learning Bars advocates the use of a demand-side model using the *EYE* data for allocating inclusive education funding. The model is being used to examine current funding allocations for First Nations students living on reserves.[56]

Professional Learning. Several studies have shown that 'one-shot' short-term workshops have little or no significant impact on student learning or teacher practice.[57] In contrast, year-long professional learning enables teachers to engage deeply with new content and instructional strategies leading to meaningful changes in classroom practice. The professional learning that accompanies the *EYE* is designed to effect changes in teaching practice and improve student outcomes.

[56] Willms, J. D. (2024). *Using the Early Years Evaluation to Estimate Special Needs Funding for First Nations On-Reserve schools.* The Learning Bar.

[57] Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. National Staff Development Council.