





Evaluating technical adequacy of DIBELS in a New Zealand sample of early elementary students Elizabeth Schaughency, Philippa Struthers, & Ruth Kaminksi schaughe@otago.ac.nz

The use of formative assessment for educational decision making in cross-cultural contexts Chantal Dufour-Martel, Chair Symposium Society for the Scientific Study of Reading Montreal, Canada 11 - 14 July 2012

Acknowledgments

The participants:

- Children
- Families

 Staff of participating schools Faculty colleagues, especially: Elaine Reese Mele Taumoepeau **Department of** Psychology Paul Thorsnes **Department** of **Economics**



Student researchers and research assistants: Abigail Pigden Helen Owen Kathryn McLennan Ryan Abraham, and especially...



Pip

Struthers

Acknowledgments – Financial Support

 Work referred to in this presentation was supported in part by a Top Achiever Doctoral Scholarship from the **Tertiary Education Commission to Philippa** Struthers, a Summer Studentship from the **Division of Sciences to** Abigail Pigden, University of Otago Research Grants to Elizabeth Schaughency, and the Department of Psychology.

 Opinions expressed herein do not necessarily reflect the position of funding agencies, and such endorsements should not be inferred.



Ko te reo te taikura o te whakaao marama

Literacy is the key to understanding

Hobbs 2003

Similarities between the literacy contexts in NZ and the US speak to roles for formative assessment tools.

Figure S.2. Student performance in reading in PISA 2006

This figure shows the percentage of students at each performance level in reading; students with scores at Level 5 are the strongest performers, those at Level 1 and below are the weakest.



Source: OECD (2007), PISA 2006, Science Competencies for Tomorrow's World, Volume 1: Analysis, Table 6.1a, available at http://dx.doi.org/10.1787/142046885031.

NZ Ranked highly in literacy attainment Documented marked variation in achievement •Similar proportions of young people reading at the highest and **lowest** levels



Potential Roles for Assessment in New Zealand

- 8 claims for enhancing literacy levels with a significant body of supporting evidence
 - These include claims that specifically relate to *assessment*:
 - Detect early
 - Maximize chances of detection for early identification of all at-risk students
 - Determine a student's strengths and weaknesses
 - Provide on-going assessment and support

Ministry of Education Literature Review: An International Perspective on Dyslexia Executive Summary



Potential Uses of Assessment for Educational Decision Making in NZ
8 claims for enhancing literacy levels with a significant body of supporting evidence
The remaining claims...
Intervene early

- Increase intensity
- Instruct in phonological awareness and phonics at an early age
- Teach focused on individual learner needs

...have implications *for using assessment information* to differentiate instruction

Ministry of Education Literature Review: An International Perspective on Dyslexia Executive Summary



Technical adequacy evidence is sample specific.¹ **EACH NEW INTENDED USE OF A MEASURE NEEDS TO BE EMPIRICALLY EVALUATED.²**

¹Christ & Hintze 2007; ²AERA, APA, NCME 1999



Contextual Differences in NZ Potentially Relevant to Assessment of Developing Literacy Skills

- Structure of beginning schooling

 5th birthday, rather than beginning of school year
- Curriculum¹
- Language²
 - Differences in vowel production patterns
 - Differences in speaking rates when reading from connected text

Small scale field trials are recommended to evaluate whether literacy measures developed elsewhere are appropriate to the NZ context.³

¹ Smith & Elley 1997
 ² Robb, Maclagan & Chen 2004
 ³ Croft et al. 2000



Results to Date • DIBELS correlate with: – School used measures and judgments^{1,2} Researcher administered criterion measures in Grade 4³ & 5⁴ -Future performance on DIBELS tasks^{1,2,3,4,5} Growth mixture models with FSF progress monitoring data suggest potential patterns of typical development and risk in kindergarten⁶

¹ Schaughency & Suggate 2007; ² Schaughency & Suggate 2008; ³ McKay, Ervin, Schaughency, Suggate & Tong 2008; ⁴Schaughency, Suggate, & Tustin, 2010, ⁵Struthers, Schaughency, Suggate, Clarke & Thurlow 2010; ⁶Schaughency, Clarke, Struthers, Beretvas in preparation



This research extends previous research by...

- Examining relations to external criterion measures to period of reading acquisition (K to Grade 2)
- Examining technical adequacy issues specifically relevant to progress monitoring, e.g.,
 - alternate forms reliability
 - predictive validity of *ideographic* (*within student*) change
- Illustrating formative assessment



Participants: The Schools

- Four primary schools in two small urban areas in NZ
 - Socioeconomic characteristics of the school communities were generally similar
 - Ministry of Education assigned deciles ranged from 3 - 4 (low - moderately low)
 - Schools varied by:
 - Demographic composition
 - School samples:
 - New Zealand European: Ranged from 56% 94%
 - Mäori: Ranged from 3% 39%
 - Participation rate
 - Ranged from 32% 76%



Participants: The Children

2009	2010	2011	-
<i>n</i> = 57	<i>n</i> = 136	<i>n</i> = 168	• Same grade samples
2000		K	year analyses (Stewart &
	K		Silberglitt, 2008).
ENS!		2	•Subsamples followed for longitudinal analyses.

US/Canadian terminolgy is used to refer to school year to communicate with a North American audience and for consistency with DIBELS manuals.



Measures:

Dynamic Indicators of Basic Early Literacy Skills ttps://dibels.uoregon.edu/ 6th edition DIBELS - Reviewed, selected, and modified for cultural appropriateness (e.g., color \rightarrow colour) - 2 rather than 3 passages administered as part of larger research protocol First Sound Fluency was substituted for Initial Sound Fluency



Data Collection

	Term 1	Term 2	Term 3	Term 4
All Participants	B1	B2 (End)		B3
Some Kindergarten Participants		FSF 2X per week	PSF/NWF 2X per week	
		With typical ar at-risk samples	nd With t and a when move Data a	ypical sample at risk sample ready to on analysis to do

Reliability Issues

- Administered by trained psychology graduate students
 - Required to achieve > 90% IRR with native speaker of NZ English
- Language measures
 - (e.g., WUF, Retell Fluency)
 - Audio recorded, transcribed, coded/analysed using Systematic Analysis of Language Transcripts (SALT) software
 - Correlations with *in vivo* scoring, e.g.,
 - WUF .93 (K, B2)
 - Retell .85 .86 (Gr 1, B2)



Reliability Issues

 Alternate form reliability - 2X/week monitoring progress of subsample. Correlations within week, e.g. • FSF .86 - .97 (K, Term 2, at-risk sample) - 2 ORF Passages administered at each data collection point. • ORF .96 (Grade 1) - .98 (Grade 2) • Retell .78 (Grade 1) - .62 (Grade 2) Average of two probes used in analyses.



Criterion Validity: WUF and PPVT4

	Concurrent	Predict	ive to
WUF assessed:		Grade 1	Grade 2
Kindergarten	.48	.17 <u>ns</u>	.37
Grade 1	.33		.41
Grade 2	.44	K B2 WUF correlat	es with exposure

to schooling and PPVT in Grade 1.

Peabody Picture Vocabulary Test - 4

- Administered Term 2 2011, raw scores
- Correlations with B2 WUF
- Partial correlations, controlling for exposure to schooling
- p < .01, unless otherwise noted



- Administered Term 4 2011 only
- Broad Reading Cluster, comprised of: Word Attack Word Identification Word Comprehension Passage Comprehension Correlations with B3 DIBELS • controlling for exposure to schooling p < .05, unless otherwise noted CRITERION VALIDITY: WOODCOCK READING MASTERY TEST – REVISED NORMATIVE UPDATE



Criterion validity: WRMT-R/NU Broad Reading

DIBELS Assessed:	Concurrent	Predict	ive to
Kindergarten	<i>n</i> = 45	Grade 1 (<i>n</i> = 63)	Grade 2 (<i>n</i> = 37)
WUF	.34	.46	.49
PSF	.62	.50	.44
LNF	.78	.69	.73
NWF	.84	.58	.74
Grade 1	n =55	\bigwedge	<i>n</i> = 32
WUF	.28		.24 <mark>ns</mark>
PSF	.27		00 <i>ns</i>
NWF	.68		.72
ORF	.80		.90
ORF Retell	.55		.68
Grade 2	<i>n</i> = 36		ative velotion of
WUF	.44	PSF holds across t	rime
ORF	.94	By end of G1 no pr	redictive
ORF Retell	.54	relation of PSF	



Findings of significant correlations extend the criterion-related validity evidence for **DIBELS** tasks during reading acquisition in NZ context **Findings of** differing relations over *time* point to need for a developmental perspective

Validity Issues: Need for Developmental Mediation Models



Social Validity Issue: School Used Measures and Judgments

Collected alongside DIBELS	Beginning	Middle	End
Kindergarten	Book Level	Book Level	Book Level School Concern
Grade 1	Book Level	Book Level	Book Level* School Concern
Grade 2*	[Book Level]	[Book Level]	[Book Level] School Concern

National Standards specify two criteria for literacy achievement for early elementary students:

- Book level targets after 1, 2, and 3 years of school
- Overall teacher judgment (above, at, below, well-below expectations)

Book level:

- The level of text at which children read with greater than 90 95% accuracy.
- Possible range: 0 30. *Some participating schools discontinued book levels after the 3 year target of 22; therefore > 22 recoded to 23.



Criterion validity: Six Year Net

Grade 1 B2 DIBELS	Book Level ^a	Observati Subt	Burt		
n = 89		Writing Vocabulary	Word Reading	Word Reading	
WUF	.42	.24	.40	.39	
PSF	.23	.15 <mark>ns</mark>	.31	.13 <i>ns</i>	
NWF	.73	.63	.65	.80	
ORF	.81	.64	.72	.88	
ORF Retell	.55	.57	.52	.52	

Six Year Net:

- Assessment after one year of school
- Administered around 6th birthday
- Partial correlations controlling for exposure to schooling p < .05 unless otherwise noted.



Predictive validity: Six Year Net

K B1 DIBELS	Book Level ^a	Observati Subt	Burt	
n = 113		Writing Vocabulary	Word Reading	Word Reading
WUF	.27	.22	.21	.31
FSF	.60	.49	.53	.54
LNF	.71	.57	.56	.66

Six Year Net:

Assessment after one year of school

Administered around 6th birthday
 Partial correlations controlling for exposure to schooling
 All p's < .05



Predicting Book Level Across Kindergarten

	Predicting Variation in Outcomes <i>between</i> Children	Predicting Variation in Outcomes <i>within</i> Children
	To what extent do Mary	To what extent does
WUF	and Joe differ in book level as a function of	Mary's book level increase as a function of
LNF	Tmportant for	an increase in
LNF + WUF	screening decisions	Important for progress monitoring decisions

- Cross-sectional time-series regression (STATA 11.0 SE)
- All p's < .01, unless otherwise noted (*p < .05)
- > Individually each task predicts variation in outcome.
- Combined LNF & WUF contribute uniquely.



Predicting Book Level Across Kindergarten

	Predicting Variation in Outcomes <i>between</i> Children	Predicting Variation in Outcomes <i>within</i> Children
	Between-estimator uses the time averages for both	Fixed effects estimator cancels out the role of other fixed
WUF	predictor and outcome ignores change over time	<i>(individual difference) factors that contribute to performance</i>
LNF	focuses on differences between cases	estimates contribution of change in predictor to change in outcome
LNF		
+ WUF	y,	$\dot{y}_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \dots, t = 1, 2, \dots$ I, where $\ddot{y}_{it} = y_{it} \cdot \overline{y}_1$ and $\ddot{x}_{it} = x_{it} - \overline{x}_i$

- Cross-sectional time-series regression (STATA 11.0 SE)
- All p's < .01, unless otherwise noted (*p < .05)
- > Individually each task predicts variation in outcome.
- Combined LNF & WUF contribute uniquely.



Predicting Book Level Across Kindergarten

	Predicting Variation in Outcomes between Children			Predicting Variation in Outcomes <i>within</i> Children		
	R ²	Slope Coefficient (SE) 95% Confidence Interval	t	R ²	Slope Coefficient (SE) 95% Confidence Interval	t
WUF	.25	.17 (.02) .1221	7.87	.31	.19 (.02) .1622	11.45
LNF	.53	.17 (.01) .1519	14.76	.47	.26 (.02) .2329	16.00
LNF + WUF	.54	.15 +19 .04*		.53	.21 + .31 .10	

Cross-sectional time-series regression (STATA 11.0 SE)

- All p's < .01, unless otherwise noted (*p < .05)
- > Individually each task predicts variation in outcome.
- Combined LNF & WUF contribute uniquely.



Predicting Book Level Beginning to Middle of Kindergarten

	Predicting Variation in Outcomes <i>between</i> Children			Predicting Variation in Outcomes <i>within</i> Children		
	R ²	Slope Coefficient (SE) 95% Confidence Interval	t	R ²	Slope Coefficient (SE) 95% Confidence Interval	t
FSF	.19	.12 (.02) .0815	6.39	.26	.17 (.03) .1222	6.47
LNF + FSF	.50	.13 + .01 <i>n</i> s		.49	.17 + .06*	

- Cross-sectional time-series regression (STATA 11.0 SE)
- p's < .01, unless otherwise noted (*p < .05; †p < .10)
- > Individually FSF predicts variation in outcome.
- Combined, results for oral language measures vary by analysis,
 - > adding to the prediction of within child change.



Predicting Book Level Beginning to Middle of Kindergarten

	Predicting Variation in Outcomes <i>between</i> Children			Predicting Variation in Outcomes <i>within</i> Children		
	R ²	Slope Coefficient (SE) 95% Confidence Interval	t	R ²	Slope Coefficient (SE) 95% Confidence Interval	t
FSF	.19	.12 (.02) .0815	6.39	.26	.17 (.03) .1222	6.47
LNF + FSF + WUF	.51	.13 + .01 <i>n</i> s + .01 <i>n</i> s		.51	.15 + .05† + .05† .20	

- Cross-sectional time-series regression (STATA 11.0 SE)
- p's < .01, unless otherwise noted (*p < .05; †p < .10)
- > Individually FSF predicts variation in outcome.
- Combined, results for oral language measures vary by analysis,
 - > adding to the prediction of within child change.



Predicting Book Level *Middle to End* of Kindergarten

	Predicting Variation in Outcomes <i>between</i> Children				Predicting Variation in Outcomes <i>within</i> Children			
	R ²	Slope Coefficient (SE) 95% Confidence Interval	t	R ²	Slope Coefficient (SE) 95% Confidence Interval	t		
PSF	.26	.16 (.02) .1220	8.07	.28	.17 (.02) .1321	8.25		
NWF	.58	.15 (.01) .1316	16.19	.31	.16 (.02) .1219	8.81		
NWF + LNF	.62	.09 + .09] .18		.41	.11 + .15			

Individually PSF and NWF predict variation in outcome.

Combined, results vary by analysis

PSF adding to the prediction of within child change.



Predicting Book Level *Middle to End* of Kindergarten

	Predicting Variation in Outcomes <i>between</i> Children				Predicting Variation in Outcomes <i>within</i> Children			
	R ²	Slope Coefficient (SE) 95% Confidence Interval	t	R ²	Slope Coefficient (SE) 95% Confidence Interval	t		
PSF	.26	.16 (.02) .1220	8.07	.28	.17 (.02) .1321	8.25		
NWF	.58	.15 (.01) .1316	16.19	.31	.16 (.02) .1219	8.81		
NWF + LNF + PSF + WUF	PSF .64	.09 + .08 doesn't add .21 .04*		.46	.08 + .10 + .29 .08 + If no PSF, WUF c	adds		

Individually PSF and NWF predict variation in outcome.

Combined, results vary by analysis

PSF adding to the prediction of within child change.



Predicting Book Level Across Grade 1

	Predicting Variation in Outcomes between Children			Predicting Variation in Outcomes within Children			
	R ²	Slope 95% Co	Coefficient (SE) onfidence Interval	t	R ²	Slope Coefficient (SE) 95% Confidence Interval	t
WUF	.35		.20 (.03) .1526	7.72	.21	.14 (.02) .1018	6.50
PSF	.17		.15 (.03) .0921	4.86	.03	.09 (.04) .0116	2.29*
NWF	.56		.11 (.01) .0913	11.86	.24	.11 (.01) .0813	7.16
NWF + WUF	.61	.09 + .09	.18		.34	.08 + .18 .10	

Individually WUF, PSF, and NWF predict variation in outcome.

Combined results vary by analysis

Either PSF or WUF add to NWF

PSF doesn't add to WUF



Predicting Book Level Across Grade 1

	Predicting Variation in Outcomes between Children				Predicting Variation in Outcomes within Children			
	R ²	Slope Coefficient (SE) 95% Confidence Interval	t	R ²	Slope Coefficient (SE) 95% Confidence Interval	t		
WUF	.35	.20 (.03) .1526	7.72	.21	.14 (.02) .1018	6.50		
PSF	.17	.15 (.03) .0921	4.86	.03	.09 (.04) .0116	2.29*		
NWF	.56	.11 (.01) .0913	11.86	.24	.11 (.01) .0813	7.16		
NWF + PSF	.59	.10 + .07		.26	.10 + .06			



Predicting Book Level *Mid to End* Grade 1

	Predicting Variation in Outcomes between Children				Predicting Variation in Outcomes within Children			
	R ²	Slope Coefficient (SE) 95% CI	t	R ²	Slope Coefficient (SE) 95% CI	t		
ORF	.68	.09 (.01) .0811	15.47	.35	.16 (.03) .0922	5.06		
Retell	.47	.18 (.02) .1522	9.93	.45	.28 (.05) .1838	5.88		
ORF + Retell	.70	.08 +13 .05		.52	.10 +31 .21			
ORF + Retell + WUF	.71	.08 + .04†16 + .04*		.55	.10 + .29 .19			
Individually ORF and Retell predict variation in outcome.								

SAFERE AUDE

Combined ORF + Retell add to prediction of increases in book level in the 2nd half of the year, results for adding WUF varying.

Today's results thus far...

• Big ideas **Big Idea 1.** Because learning and development imply change over time, what is evidence based assessment will also change. **Big Idea 2.** Different types of evidence are needed to support different purposes of assessment.



- context, predicting both between and within children over time
- 'Within' analyses were generally more sensitive to detecting unique contributions of • measures with an oral language component
- Sensitivity of predictor and outcome measures may influence obtained findings.

Validity Issues: Need for Developmental Mediation Models





Growth mixture modelling suggests 2 developmental patterns for phonological awareness skills across Term 2.

- The typical pattern, shown above, shows varying skill levels at the beginning of the term, with convergence over the term.
- The intercept for the second pattern is considerably lower, with limited growth across the term, potentially suggesting developmental risk.
- Schaughency , Clarke, Struthers & Beretvas in progress

Sample Progress Monitoring Results

2 students with BL = 1

Tipene Baseline Progress Monitoring Term 1







Figure 3: First Sound Fluency





Findings of change in trend with supplemental support suggests sensitivity of FSF to treatment effects. **Book level** findings speak to need to also consider literacy skills necessary to apply PA to reading (Case 2)

The Moving Target of Literacy Development



- formative assessment measures in New Zealand, issues such as:
 - Appropriate cut-scores and indices for decision-making
 - Factors related to uptake & instructional validity

Thank You



 American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Christ, T. J., & Hintze, J. M. (2007). Psychometric considerations when evaluating response to intervention. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.). Handbook of response to intervention: The science and practice of assessment and intervention (pp. 93-105). New York: Springer.



 Croft, C., Strafford, E. & Mapa, L. (2000). Stocktake/evaluation of existing diagnostic tools in literacy and numeracy, in English: A report to the Ministry of Education. Wellington, NZ: New Zealand Council for Educational Research. Kaminski, R., Cummings, K. D., Powell-Smith, K. A., & Good, R. H. (2008). Best practices in using dynamic indicators of basic early literacy skills for formative assessment and evaluation. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology V (Vol. 4) (pp. 1181 – 1204). Bethesda, MD, US: National Association of School Psychologists.

- McKay, L. D., Ervin, R. A., Schaughency, E., Suggate, S. & Tong, J. (2008, June). Evaluation of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) for Predicting Literacy Attainment in Canada and New Zealand. Paper presented at the 69th Annual Convention of the Canadian Psychological Association, Halifax, Nova Scotia, Canada, 12 – 14, June.
- Ministry of Education (2006). Better outcomes for children: An action plan for GSE 2006-2011. Wellington: Ministry of Education.



• Ministry of Education (n.d.) *Literature* review: An international perspective on dyslexia. Retrieved from http://www.tki.org.nz/r/literacy_numeracy • Ministry of Education (2009). The New Zealand Curriculum Reading and Writing Standards for Years 1 - 8. Wellington: Learning Media.



- OECD (2007). PISA 2006, Science competencies for tomorrow's world. Vol. 1: Analysis
- Robb, M. P., MacLagan, M. A., & Chen, Y. (2004). Speaking rates of American and New Zealand varieties of English. *Clinical Linguistics & Phonetics*, 18, 1 – 15.



 Schaughency, E. & Suggate, S. (2007, August). Evaluation of DIBELS for Assessing Literacy Growth in New Zealand. Paper to be presented at the 115th Annual Conference of the American Psychological Association, San Francisco, CA.

 Schaughency, E. & Suggate, S. (2008). Measuring basic early literacy skills amongst Year 1 students in New Zealand. New Zealand Journal of Educational Studies, 43(1), 85 – 106.



- Schaughency, E., Suggate, S. & Tustin, K. (2010). Using local norms to describe reading achievement. In E. Schaughency, Chair, *Different questions, different queries: Looking at literacy development in a new light.* Symposium presented at the American Educational Research Association 2010 Annual Meeting. Denver, CO.
- Smith, J. W. A. & Elley, W. B. (1997). How children learn to read: Insights from the New Zealand experience. Auckland, NZ: Longman



• Struthers, P., Schaughency, E., Suggate, Clarke, & Thurlow (2010). Using a back-mapping framework to examine early literacy skills in the first year of school. In E. Schaughency, Chair, Different questions, different queries: Looking at literacy development in a new light. Symposium presented at the American Educational Research Association 2010 Annual Meeting. Denver, CO. • Woolridge, J. M. (2003). *Introductory* econometrics: A modern approach. Mason, OH: Thomson South-Western.

