

A comparison of the effects of text-based instruction versus standards-based instruction in the early years



Authors:

Bonnie Grossen¹
Nithi Muthukrishna²
Jaqueline T. Naidoo²

Affiliations:

¹Centre for Applied Research
in Education, Portland,
United States

²School of Education,
University of KwaZulu-Natal,
South Africa

Corresponding author:

Nithi Muthukrishna,
muthukri@ukzn.ac.za

Dates:

Received: 20 May 2016
Accepted: 15 June 2016
Published: 29 Nov. 2016

How to cite this article:

Grossen, B., Muthukrishna,
N. & Naidoo, J.T., 2016, 'A
comparison of the effects of
text-based instruction versus
standards-based instruction
in the early years', *South
African Journal of Childhood
Education* 6(2), a459. [http://
dx.doi.org/10.4102/sajce.
v6i2.459](http://dx.doi.org/10.4102/sajce.v6i2.459)

Copyright:

© 2016. The Authors.
Licensee: AOSIS. This work
is licensed under the
Creative Commons
Attribution License.

The purpose of the study was to evaluate the effectiveness of a text-based (TB) English structured language development programme and a teacher-designed standards-based (SB) English instructional model. The sample of this study comprised of 500 Samoan children, in Kindergarten (K) and Grade 1 (G1), on the island of American Samoa attending eight different schools. All the children enter school with no English competence. Six schools implemented the TB Language for Learning scripted programme and the Read Well. Two schools implemented the SB instructional model for English language development (SB), and only one of these schools implemented the SB instruction in reading. The results of this study support the efficacy of TB structured language programme as compared to the teacher-designed SB instructional model in all language and reading skills assessed.

Introduction

Internationally, in countries of the North and South, there is a concern regarding effective and sound instructional practices for English language learners and the most effective way to enable these learners to access and make significant progress in the grade level academic curriculum (Brock-Utne 2007; Rossell 2005; Taylor & Coetzee 2013). These debates have raised issues of ideology, values, social justice and equity, pedagogy and the very core of schooling for English language learners in socially, economically, culturally and linguistically diverse societies. Teaching English to learners who speak languages other than English is the focus of our paper in this special issue. English may either be the dominant language of a country or a global language of wider communication used as the medium of instruction in schools.

Globally, there have been various approaches to support learners who speak languages other than English. Advocates of instruction in the *mother tongue* argue that while children are learning to speak English, they should be taught to read and write in their mother tongue first (for example, Brock-Utne 2007; Hakuta, Butler & Witt 2000). Further, failure is likely if children are required to learn both oral English and English reading and writing at the same time. In this model, students are transitioned to English-only instruction when their oral English is adequate to ensure success, usually in the third or fourth grade. *Structured English Immersion (SEI)* programmes have also been the subject of much debate in research. The view is that given sound high-quality programmes, learners can acquire content knowledge, concepts and skills at the same time that they improve their English language skills through direct instruction (Hayes, Salazar & Vuckovic 2002). SEI programmes involve active, direct and explicit teaching methods and strategies, and teachers use and teach English at a level appropriate to the abilities of the learner (Clark 2009). Proponents of SEI hold that through the study of grade-level curriculum content, students engage with meaningful topics and material that are relevant to their academic programmes (Clark 2009). The aim is to maximise instruction in English, and more important, language instruction is not separate from the other academic content areas of the curriculum. SEI involves intensive, carefully structured support through the use of modelling, demonstrations, graphic organisers, adapted texts and visual aids. Teaching aims to link the language processes of listening, speaking, reading and writing so that these develop simultaneously and interdependently. Furthermore, the assumption is that English is learned best when learners interact often during the school day in meaningful contexts (Rossell 2003). Gersten and his team of researchers found SEI to produce better learning outcomes than bilingual education with Vietnamese immigrants in California and Hispanics in Texas (Becker & Gersten 1982; Gersten 1985; Gersten & Woodward 1985; Gersten & Woodward 1995). Studies in Canada and Arizona, USA, have found SEI programmes to be highly successful and effective (Arizona Department of Education 2004; Genesee 1985). Opponents of immersion programmes have claimed that learners are likely to fall behind in their academic progress as the immediate focus is on English language development (Sievert 2007). Further,

Read online:



Scan this QR
code with your
smart phone or
mobile device
to read online.

concerns have been voiced that children are likely to lose their mother tongue in the process of developing English.

Consequently, in schools in the majority of international contexts *bilingual education* is offered to English language learners who speak languages other than English (Cummins & Hornberger 2008). Although bilingual education is often poorly defined in practice and there are vast disparities in how it occurs in practice, it entails an organised and planned programme that uses two languages. These programmes range from those that aim to develop equal proficiency in the two languages to those that do not have such a goal. For example, in a transitional bilingualism paradigm education, adopted in the school curriculum in most of sub-Saharan Africa, the aim is for learners to first acquire language fluency and access the academic curriculum in the mother tongue, and when learners are able to access instruction in English, the home language instruction is phased out, and learners are taught exclusively in English (Cummins 2009). Critics of bilingual education claim that programmes incur significant costs but there is limited long-term data on the effectiveness of bilingual education (Sievrt 2007). Faltis (2011) asserts that empirical studies reflect mixed results on the effectiveness of bilingual programmes in education.

As a further alternative, *immersion models* involve the immersion of learners in a classroom environment where teaching and learning is conducted almost exclusively through the second (or third) language. The argument is that in schools where English is the medium of instruction, learners have the best chance of academic success if they are able to access curriculum content, concepts and skills at grade level as early as possible (Cummins 2009). These programmes vary in the degree of support provided which can range from extensive support provided by well-trained language teachers and in other cases learning is mere assimilation which results in a 'sink or swim' experience (Cummins 2009). The concern for education policy makers has been about how best to preserve the mother tongue while preparing its future citizens to function in and participate in the larger world community (Huebner 1986).

We argue in this paper that the issue of models of English language teaching and the language of instruction are critical issues but equally important. Educational researchers and teachers need to interrogate the quality of instruction offered, that is, the instructional models, programmes and strategies. Currently in most schools in South Africa, for example, the first language is the dominant language in grades 1, 2 and 3, and learners transition to English as the language of instruction in grade 4. Further according to the new Curriculum and Assessment Policy Statements (CAPS), English is introduced in Grade 1 (G1) in all schools (Department of Education 2011a). However, since the democratic government came into power in 1994, it is a grave concern that South Africa is a long way off from achieving the ideal of excellence in education.

South African children perform poorly on international, regional tests and national assessments of educational achievement. South Africa was one of the forty countries that participated in the *Progress in International Reading Literacy Study* (PIRLS) in 2006. Approximately 30 000 grade 4 and 5 learners were assessed. Alarming findings were reported, in particular, grade 4 and 5 learners achieved the lowest scores compared to other participating countries. South Africa's Annual National Assessments (ANA), involving almost six million learners in primary schools in grades 3 and 6, showed that despite the interventions by the Department of Education, results in literacy were disappointingly low. The best provincial figure for grade 3 literacy was 46% in Western Cape. The average percentage scores for literacy in grade 3 were 36%; and in grade 6 the figure was recorded at 28% (Department of Education 2011b).

Critics place the blame for poor educational outcomes in South African schools squarely on the legacy of Curriculum 2005 and its outcomes-based approach adopted in March 1997. Critics blame the ideology of learner centredness, standards/outcomes-based approaches and progressive pedagogy thrust upon an education system with under-resourced schooling contexts, poorly qualified teachers, teachers with poor subject knowledge, lack of learning support materials and poor quality teacher development programmes. To address poor achievement outcomes in education, in 2008 Naledi Pandor, Minister of Education, launched the *Foundation for Learning Campaign*, which focused on the Foundation and Intermediate phases of education (Department of Education 2008). A more text-based (TB) approach was adopted with curriculum materials reflecting a more explicit pedagogy, clear purpose, pacing, sequencing, coherent set of targets, time allocations, activities and resources required to improve reading, speaking, listening and writing. The impact of this programme is still to be seen.

In the present study, our aim was to compare the effects of a TB structured immersion language development programme with the effects of a teacher-designed standards-based (SB) instructional model on English language development and reading. The TB model used in the present study is the same model evaluated in a study in South Africa, in a comparison with a child-centred model of instruction (Grossen & Kelly 1992). In that study, an indigenous teacher presented the TB instruction to 6 year-old children. The English proficiency of the TB group was compared with that of the children the same teacher had taught the year before. Those children were, of course, one grade higher. On all measures of English reading and language, the TB group scored significantly higher than the group taught by the same teacher the previous year and significantly higher than the children 1 year older in a different school in the same homeland.

Grossen and Kelly (1992) reported a second comparison. The English proficiency of the TB group, when they were 7 years old, was compared with that of same-age children taught in a similar school (control), and in a more advantaged multiracial school, taught by a native English-speaking teacher. The TB

group scored significantly higher in both English reading (word recognition and comprehension) and English language (receptive and expressive) than the control group. The differences in performance of the TB group and the multiracial group were non-significant, except for receptive language, where the multiracial group scored higher. This comparability in performance is a remarkable achievement for the TB group because of its disadvantages. The TB group size ranged from 45 to 55 children during the instruction, while only 19 children were enrolled in the multiracial class. The TB group was taught by an indigenous teacher, while the multiracial group was taught by a native English-speaking teacher. In spite of the disadvantages of a doubly large instructional group, and a far less qualified teacher, the TB group's performance was quite comparable to that of the multiracial group.

In the present study, our sample comes from the population of American Samoa in the South Pacific. According to education policy, the Samoan language is of equal importance with English. English is the medium of instruction but a period of Samoan language instruction is also included in the curriculum. American Samoa has much in common with other indigenous societies: High levels of poverty, family fragility and dysfunction, and social isolation. As in South Africa, the education system of American Samoa faces challenging problems: A lack of highly-qualified teachers, inadequate facilities and children who come to school with little or no competence in English.

We present a large-scale study of the effects of the TB model of English language instruction on the English language development of children aged 5 and 6 in American Samoa. Our control group are children being taught with the new SB model of instruction being implemented there. The SB model is the same model as the outcomes-based model of instruction promoted recently in South Africa.

Sample

American Samoa provides public education in 22 elementary schools. The school year runs from mid-August to mid-June. Each elementary school includes a Kindergarten (K) for children age 5 or turning 5 by December 30. Children age 6 or turning 6 before December 30 are placed in G1. Each elementary school serves K and Grades 1 through 8.

Although the biggest island of American Samoa is only 16 miles long and 5 miles wide at the largest measure, 12 of the schools are considered 'remote' and serve only very small populations of children (< 20 in G1). For example, travel to the nearest remote school would require more than an hour by car from the central office location. Initially, only two remote schools were included in the sample to reduce the expense of monitoring and data collection. The 10 larger schools were located in the more densely populated areas of the island.

All 10 larger schools were initially included in the sample. After implementation, four of the larger schools were

eliminated due to chronic teacher absence in both grades K and 1. Half the teachers in these schools were absent for more than half the year. No substitute teacher system was available to the schools, so when teachers were absent, children of the missing teachers were placed in classrooms with other teachers of that level, resulting in sometimes 60 or more children in one classroom. When teachers are chronically absent, instructional quality is seriously compromised.

Three more of the larger schools were eliminated from the K language study due to implementation failure. Managers in charge of the language implementations were absent for most of the year. One additional small school was eliminated from the reading study due to chronic teacher absence in G1. At the time of the evaluation, subjects were 500 Samoan-speaking children in K and G1 attending eight different public schools.

The children came from a very homogeneous culture where Samoan language is spoken exclusively in the homes and in the community. (The five children who came from English-speaking homes were excluded from the study.) All the children enter school with no English competence. The schools they attend are under the jurisdiction of the United States and the medium of instruction is English.

As a matter of policy, English is the medium of instruction throughout the school day, except for the period of Samoan language instruction. However, 99% of the children come to school speaking no English at all, and the teachers themselves are more fluent in Samoan than in English. For these reasons, in real practice Samoan tends to be the main medium of instruction in the early grades.

Acquiring English competence is one of the biggest challenges of the educational programme of American Samoa. New children who have had no schooling constantly arrive from neighbouring islands (Western Samoa, Tonga). It is not uncommon to find Grade 4 classrooms where the children still speak no English, read no English sounds and, generally, have no English competence. The teachers' English competence is also often limited. One often hears teachers making frequent grammatical errors in their speech, especially in subject-verb agreement, singular and plural distinctions and any other English inflectional endings.

Two instructional models

Text-based instruction

The teaching materials in the TB treatment included a language-development programme and a reading programme. The language development programme *Language for Learning* (Engelmann & Osborn 2008) was highly structured and scripted for the teacher. Through this oral programme young children are able to acquire the knowledge and understanding of words, concepts and statements they would need for a beginning English reading and writing programme.

Following the script of the language development programme, the teacher presented a rapid succession of directions and tasks that required the children to respond as a group using English. The teacher gave immediate corrective feedback when errors occurred. After the group of children appeared to be at mastery on a set of tasks, the teacher gave the children individual turns to check for mastery. Sets of tasks were repeated until every child was at mastery. On subsequent days, the teacher presented a different set of the same type of tasks, testing understanding of the same concept. When the group of children gave the correct response the first time, without requiring correction, the concept was considered mastered. Usually first time success occurred on the 3rd day of practice. Each daily lesson required work on multiple concepts. For example, labelling nouns ('A dog'), saying the whole sentence ('That is a dog'), following directions ('Touch your nose'), asking and answering personal questions ('What is your name?' 'What is his name?'). Teachers received a half day of instruction in how to follow the programme script, followed up with occasional side-by-side coaching and after-school feedback sessions.

The TB reading programme for K and G1 was *Read Well* (Voyager Sopris Learning 2004). The *Read Well* programme specifies details of the instruction, including a teaching script, sequence of reading tasks and learning activities, along with a storybook and workbook pages. The *Read Well K* programme was used in K both during the current school year and the previous school year. The 38-unit *Read Well 1* programme was used in G1. G1 children had received both the *Read Well K* instruction the previous year, and *Read Well 1* during G1. Most children completed 20 to 25 of the 38 units in G1.

In K, the TB language instruction was presented during a 30-minutes block and reading in another 30-minutes block. In G1, reading was expanded to a 60-minutes block. All the instruction in both language and reading was presented in English language. However, when children were confused, the teacher would translate into Samoan. During another 60-minutes period of the school day, the children received instruction in Samoan. They learned songs in Samoan, recited rhymes and listened to traditional stories. During the remainder of the school day, the children would hear both Samoan and English. The teacher would use English as much as possible, but also translate the same information into Samoan.

Standards-based instruction

The teacher-designed SB instruction required teachers to find or develop lessons that taught the concepts outlined in a set of standards (or outcomes) developed for American Samoa. Teachers were restricted from using a textbook to guide their instruction. Instead they created detailed lesson plans to teach the learning objective specified in a standard for that grade level and used multiple resources. Teachers received in-service training in how to develop SB lesson plans in 2 days before school started plus in after-school meetings during the year. Experts reviewed their lesson plans once a month and gave feedback. The teachers had

shared lesson plans in the first year of implementation, with one teacher writing the plan for all G1 teachers for reading, and another writing for mathematics, and so on, to reduce their workload. The SB model experts asked the teachers not to share, but to each write their own lesson plan for all subjects. So each of the seven G1 teachers wrote their own unique lesson plans during the year of the study. The teachers did have the *Language for Learning* programme and *Trophies* (Harcourt 2007) programme available in their classrooms, but they were expected to use these programmes as a resource as they selected learning activities that aligned with the standards of American Samoa. They were not allowed to follow the instructional sequence of the programme.

The standards for K included the following performance indicators for speaking and listening (English language):

- Name people, places, things and events, and describe them with appropriate details.
- Recount an experience and express thoughts and feelings, speaking audibly.
- Repeat a word or phrase (e.g. 'reality', 'tomorrow' and 'come with me') using different kinds of expression (e.g. with excitement, surprise, joy, suspicion, urgency and secrecy) and levels of intensity (whisper, shout) and use gestures to accompany the message.
- Identify by locating and repeating the names of objects or print words stated orally (e.g. in a text, on a bulletin board, in a poster).
- Sort common objects into categories (e.g. shapes, foods) to gain a sense of the concepts the categories represent.
- Use words and phrases acquired through conversations, reading and being read to, and responding to texts (e.g. new nouns and verbs, descriptive words and sensory details).
- Identify real-life connections between words and their use (e.g. identify places at school that are colourful, busy and quiet).
- Use picture dictionaries to determine and clarify the meanings of words.
- The standards for reading for G1 included the following performance indicators.
- Distinguish long and short vowel sounds in spoken single-syllable words.
- Count and blend syllables in spoken words.
- Add or substitute individual sounds (phonemes) in simple, one-syllable words to make words.
- Decode regularly spelled one-syllable words.
- Read appropriate level text with purpose and understanding.
- Read elementary text orally with developing accuracy and rate/speed on successive readings.
- During the 60-minutes English language arts period, the SB teachers used English as the medium of instruction. The SB teachers also translated into Samoan any unclear information they had presented in English. The rest of the school day, including the 60-minutes Samoan language period, was the same for both groups.

Both instructional models received frequent coaching during the year to ensure implementation fidelity through support and coaching feedback. Classroom observations that included feedback and after-school meetings for both treatments occurred on a consistent, uniform schedule. The schedule required programme implementation specialists to observe each teacher about once every 2 weeks. The programme specialists were specifically trained for each model. The SB-treatment was the new model and received extensive support from the Regional Service District based in Honolulu, which was promoting SB instruction. The study occurred in the second year of the implementation of the SB model.

Measures

All the measures were individually administered. Independent evaluators, consisting of eight island-based external support personnel and four off-island specialists, individually tested all the children in grades K and G1. The school programme required student participation in the instruction and assessment as part of the regular curriculum and evaluation procedures. Thus, the research qualified for exemption from informed consent. The data had no personal identifiers by student, and the research presented 'no more than minimal risk of harm to subjects' (Resnik 2005:36).

Prior to instruction, the English competence of each K child entering school was assessed with a very simple, quick measure. As each child was asked to say his/her name for matching to the protocol, the tester scored the child's response (see box 1).

This scoring gave a quick measure of the level of the child's English competence. A score of 4 represents a minimal level of English competence: the children can answer the question 'What is your name?' with their name.

We used two tests to evaluate the effectiveness of the language development treatments. The first test was a Test of Beginning English Acquisition (TBEA) that was developed by Grossen and Kelly (1992) to evaluate the English acquisition of South African children learning English as a second language.

Grossen and Kelly used Cronbach's coefficient alpha statistic to calculate internal consistency reliability. Cronbach's coefficient alpha is a statistic derived from the pairwise correlations between items. Internal consistency scores range between negative infinity and one. Cronbach's coefficient alpha will be negative whenever there is greater within-subject variability than between-subject variability. An alpha of greater than + 0.7 indicates a measure with acceptable

internal consistency. Cronbach's coefficient alpha for the overall TBEA test of 39 items was 0.95. The TBEA consisted of two parts, receptive and expressive language. The receptive test (Cronbach's coefficient alpha = 0.88) had two pictures. For the first picture the children heard statements describing the activities, appearance, or position of depicted cats and were required to point to the described cat. The second picture showed various animals in a barnyard. The children pointed to the animals as they were named by the tester. The expressive language portion of the test (Cronbach's coefficient alpha = 0.72) required pupils to respond to a series of directions, such as 'hold your hand behind your head', and answer questions about what they 'are doing' and what they 'were doing'.

The second test was the vocabulary section of the Preschool Language Scale – 5th Edition (PLS-5) (Zimmerman, Steiner & Pond 2011). The test was shortened by using only the first 34 vocabulary items. The child was presented with pictures and asked 'What is this?' The items to be labelled included cookies, refrigerator, spoon, elephant, fish, ball, bird, balloons, shoes, tricycle, glass, pan, TV. For other pictures, the child was asked, 'What is this boy doing?' The expected answers were playing, eating, sleeping, drinking, running, washing or bathing.

The PLS-5 is designed to measure language growth of English-speaking children from birth to age 7. By deleting the later items, which hardly any children would be expected to know, the amount of time required to administer the test was greatly reduced (no child got all the items right so the shorter test did not have a ceiling effect). By using two short tests, we could use one test to validate the other, and take less time to administer the test. The combined test required approximately 10 to 15 minutes to administer to each child.

We assessed the emergent reading skills of the K children using the *Dynamic Indicators of Early Literacy* measures of Correct Letter Sounds (CLS) and Whole Words Read (WWR) (*DIBELSnext*; Good & Kaminski 2011). In addition to administering the CLS/WWR, we also administered the Oral Reading Fluency (ORF) to the G1 children. The *Dibelsnext* Technical Report (Good *et al.* 2011) reports the following important statistics regarding the technical adequacy of the measures. For the CLS and WWR measures, the inter-rater reliability was 0.99 for each measure in both K and G1. The *DIBELSnext* measure ORF that was used for G 1, consists of a set of three 1-minute timed readings of three different validated grade-level passages for each student. The inter-rater reliability of the ORF was 0.99. Two measures were taken from the ORF, the number of words read correctly in 1 minute Words Correct Per Minute (WCPM) and the percent accuracy (number of correct words divided by total words read). The alternate form reliability of the three forms for WCPM was 0.98; and for accuracy, 0.88. These reliability values are high, indicating that the measures are of high quality.

BOX 1: Ask in English, 'What is your name?'

Score Response

- 4 Child responds immediately with name. _____
- 3 Child responds with prompting/repeating question in English. _____
- 2 (Ask in Samoan, What's your name?)
Child does not seem to understand English. (Answers when asked in Samoan) _____
- 1 When asked in Samoan, child still does NOT give name. _____

English language competence study

Method

Schools were not randomly assigned to the TB and SB treatments. The schools for the SB treatment were selected for their proximity to the airport and the central office. The SB implementation was a pilot monitored closely by the experts from Honolulu. The pilot was in its second year. The remaining schools were in the TB treatment, which had been implemented for 7 years.

Prior to instruction, K children entering school were evaluated using the 'What's Your Name?' measure. Most of the K children in the TB *Language for Learning* instructional groups covered 40 to 80 lessons of the 120 lessons in the programme and completed the *Read Well K* programme. The SB treatment completed all the prescribed standards.

In March/April of the school year, the two measures of language development (the TBEA and PLS-5) were administered to every other K child to evaluate the effectiveness of the two instructional models, TB and SB. Because the tests were individually administered, evaluating only half the children reduced costs and allowed for adequate sample size.

Statistical significance between the two treatment groups was calculated using *t*-tests. We used the Bonferroni method to correct for familywise error (false positives) when doing multiple comparisons (Dunn 1961). Using the traditional value of 0.05 to define the significance threshold, we divided 0.05 by the number of comparisons, 13, to redefine the threshold for significance for this study. That revised threshold was 0.004. The *t*-test of each comparison must yield a probability value less than 0.004 for the family of multiple comparisons to be considered statistically significant at the 0.05 level.

The effect size (ES) is a measure of educational significance (Cohen 1969). With larger samples, smaller differences between groups can reach statistical significance but lose practical significance. ES is a measure of the difference

between groups in standard deviations (divide the difference in the mean scores of the two groups by the pooled standard deviation). Cohen (1969) describes an ES of 0.2 as small, 0.5 as medium and 0.8 as large in terms of educational significance. Cohen equates an ES of 0.8 to the difference between a college freshman and a person with a PhD (the differences between the two treatment means is far greater than the variability of scores within each treatment).

Results

At the end of 7 months of treatment, only five schools were included in the K evaluation. Schools were eliminated because a) no program specialist visited the school to provide materials and coaching (two schools), and b) chronic teacher absence (one school). Table 1 displays the results of the English competence test for the five schools remaining in the study on the 'What's your name?' test. A higher percentage of children scoring '4' and a lower percentage scoring '2' would indicate a higher level of English competence in that cohort.

The largest school in the TB treatment included the most children scoring 2 (13%), indicating no English competence at all. These are children who could not answer the question 'What's your name?' in English but did immediately answer when asked in Samoan. This school was also the largest school on the island, with a total of over 12,000 children in grades K to 8. The largest school in the SB treatment had 7% of the K children scoring 2 at the beginning of the year. Overall, for the TB treatment, 79% of the children scored '4', and 7% scored '2'; for the SB treatment, 85% scored '4', and 6% scored '2'. These scores show a slight bias favouring the SB group prior to treatment.

Table 2 displays the results of the analysis of the performance of the two treatment groups of children using the mean combined total score, the mean score on the PLS-5, and the mean score on the TBEA. Because both measures included a receptive and expressive component, these dimensions were also evaluated. The mean receptive score in Table 2 represents the total number of receptive language items correct on both the PLS-5 and the TBEA (total number of items = 26).

TABLE 1: Percentage of entering K students receiving each score by school. Schools are ordered from greatest English competence to least competence, within each treatment group.

Treatment group	Score 4 (%)	Score 3 (%)	Score 2 (%)	Score 1 (%)
TB-M, <i>n</i> = 41	98	2	0	0
TB-L, <i>n</i> = 69	83	6	4	7
TB-P, <i>n</i> = 86	67	6	13	14
SB-M, <i>n</i> = 98	85	2	7	6
SB-S, <i>n</i> = 14	79	14	0	7
Total, <i>n</i> = 308	-	-	-	-

n, students tested.

TABLE 2: Comparison of mean scores on English language competence measures.

Treatment group	<i>n</i>	Total (SD)	PLS-5	TBEA	Receptive	Expressive
Text-based	80	43.9 (12.0)	23.3 (5.8)	20.6 (7.1)	16.7 (3.6)	27.1 (9.3)
Standards-based	53	25.9 (12.0)	15.3 (6.5)	10.6 (6.8)	11.0 (4.8)	14.9 (8.2)
Effect size	-	1.50	1.29	1.45	1.38	1.39
Probability	-	<i>t</i> = 8.5; <i>p</i> < 0.0001	<i>t</i> = 7.4; <i>p</i> < 0.0001	<i>t</i> = 8.1; <i>p</i> < 0.0001	<i>t</i> = 7.8; <i>p</i> < 0.0001	<i>t</i> = 7.8; <i>p</i> < 0.0001

Similarly, the expressive language mean scores in Table 2 includes all expressive language items from both measures (total number of items = 54).

The probability of differences in the means occurring by chance exceeded the 0.004 threshold and were extremely significant at the $p < 0.0001$ level. So clearly the TB *Language for Learning* programme had significantly more favourable effects on the English language acquisition of the K children in American Samoa than did the teacher-designed, SB instruction.

Figure 1 displays the differences in mean raw scores of each treatment group graphically by school. Each of the TB treatment groups in each school achieved an average score higher than the highest average raw score of the SB schools. The lack of variability across schools indicates a consistently positive effect for the TB treatment.

Emergent reading

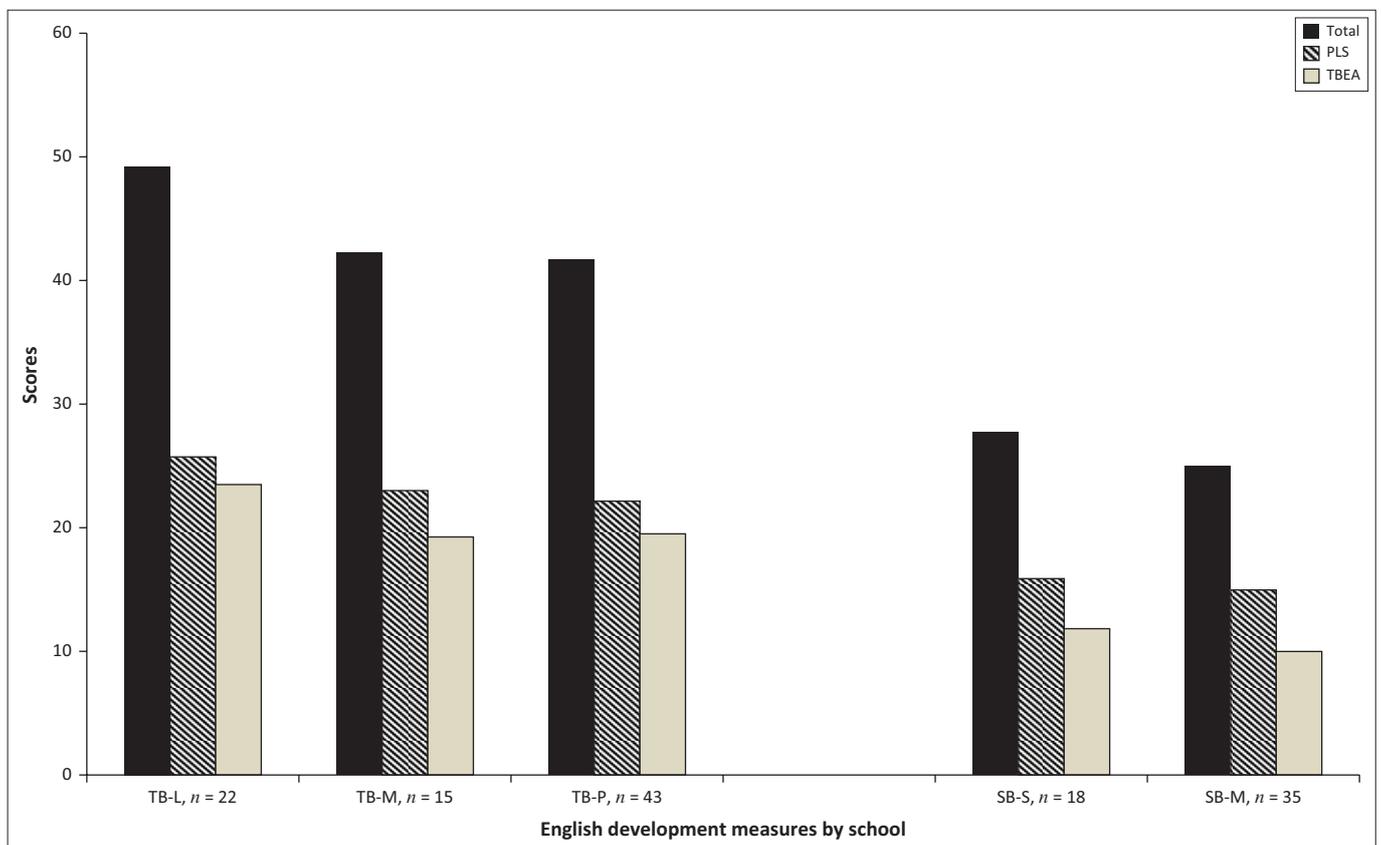
Method

Of the eight schools originally included in the overall study, one small school was eliminated from the G1 SB treatment because of teacher shortage. G1 children were taught with grades 2 and 3 children, in a group of 50–60 children. In the six TB schools The *Read Well K* and *Read Well 1* programmes were implemented for 2 years. Children in G1 had received the TB instruction in K the previous year, and in G1. The SB

programme also used the SB model for the previous year in K and in the current year. We evaluated the children's emergent literacy in K using the *DIBELSnext* measures of CLS/WWR. In G1 we used the CLS and the ORF, which included a measure of WCPM and % Accuracy. At this point the G1 children had received nearly 2 years of the treatment instruction. We had administered the CLS measure to the G1 children at the end of K the previous year. The CLS measure was the best measure (with adequate variability at the end of K and at the end of G1) to allow a valid measure of growth from end of K to end of G1.

Results

Six schools implemented the TB instruction in reading in K and then in G1 ($n = 361$). One school ($n = 103$) implemented the SB instruction. In March/April, near the end of the year, the CLS and WWR measures from the *DIBELSnext* were administered to the K children in the TB and SB treatments. Table 3 shows the differences in the performance of the K children learning in these respective treatments. The difference in the means on both the CLS and WWR measures was statistically significant (the difference on the WWR measure just met the threshold for significance). The effect sizes of 0.5 and 0.4 were educationally significant. Figure 2 displays the results by school. Each of the TB treatment groups in each school achieved average scores that were higher than the average score of the SB treatment on both the CLS and WWR measures. Note again the consistency of



TB, text-based; SB, standards-based.

FIGURE 1: Raw mean scores of K children on English development measures by school, TB versus SB.

better achievement for the children in the TB treatments over the achievement of the SB school graphically displayed in Figure 2.

Growth in emergent literacy was evaluated from the end of K to the end of G1 using the *DIBELSnext* measure of CLS. We administered the CLS measure to the children in G1 at the end of K the previous year and at the end of G1. Figure 3 displays in stacked bars the average number of sounds children read correctly for letters at the end K (the grey part of the bar) and the average score for the same children at the end of G1 on the same measure. The black part of the bar shows the growth from end of K to end of G1. We included only children with a score both at the end of K the previous year, and at the end of G1 in this analysis. The children learning from TB instruction in all five schools achieved consistently higher scores at the end of G1 and achieved consistently greater gains than the children learning from the SB instructional model. Table 4 displays the statistics for these treatment comparisons. ES of 0.50, 0.73 and 0.67 show

TABLE 3: Comparison of mean scores of K children on the *DIBELSnext* measures of Correct Letter Sounds and Whole Words Read.

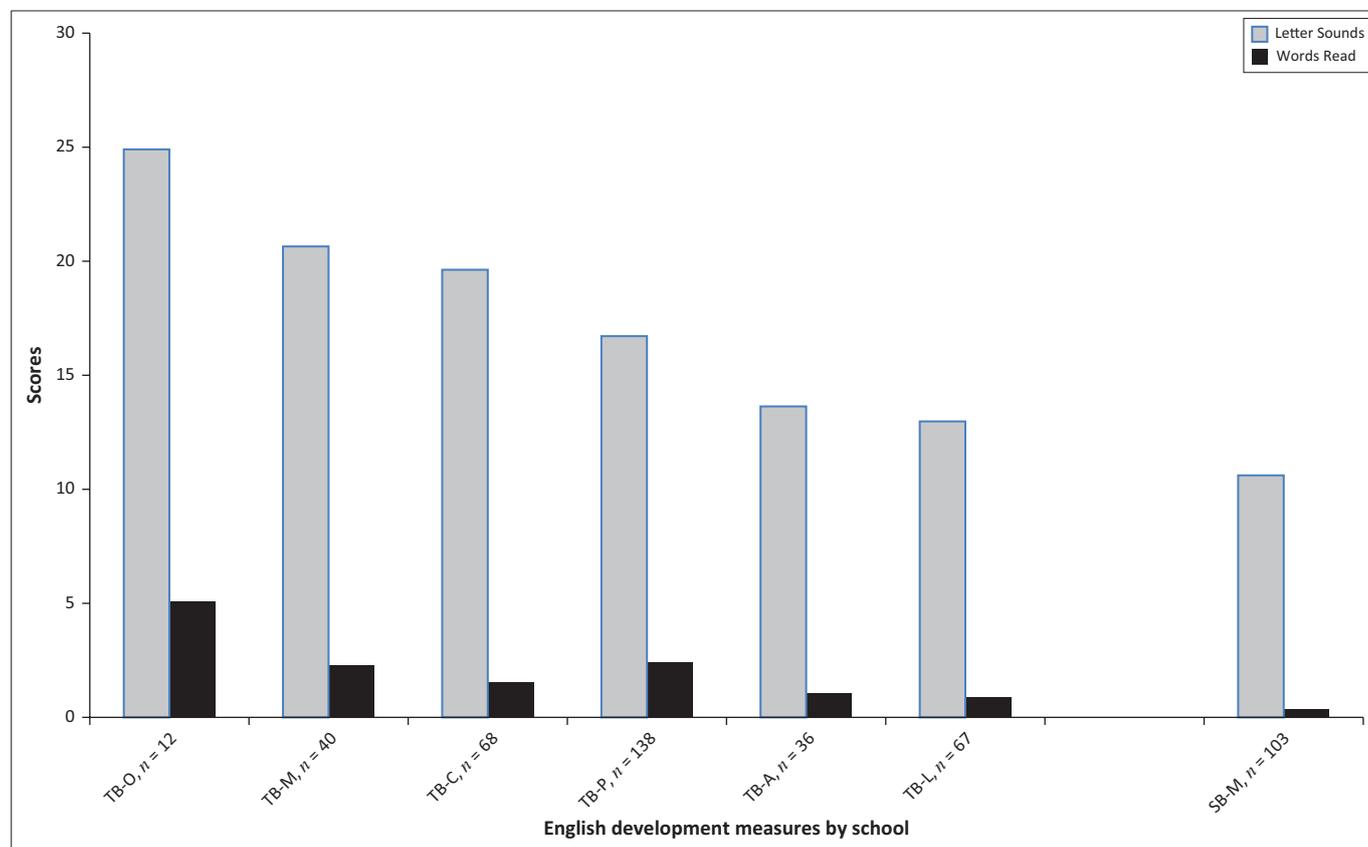
Treatment group	<i>n</i>	CLS	WWR
TB	361	17.0 (16.1)	1.9 (5.3)
SB	103	10.6 (9.3)	0.35 (2.3)
Effect size	-	0.5	0.4
Probability	-	$t = 3.5, p < 0.0001$	$t = 2.9, p = 0.004^*$

*. Just meets the redefined significance threshold of $p = 0.004$ to correct for multiple comparisons.

strong practical significance favouring the TB treatment. The differences between the means of the TB and SB groups and the difference in the mean change from end of K to end of G1 also achieved strong statistical significance at a $p < 0.0001$ level, favouring the TB treatment (see Table 4). (Figure 3 and Table 4 included only matched scores.)

The ORF measure was also administered at the end of G1. Table 5 displays the average scores on the ORF as well as the CLS measure. Table 5 also displays the results of the statistical analyses of the treatment differences. Effect sizes of 0.87, 0.60 and 0.70 show great educational significance favouring the TB treatment. The *t*-tests showed very strong statistical differences in the means ($p < 0.0001$), well within the threshold for multiple comparisons of $p < 0.004$. Figure 4 displays these results by school for visual inspection. The by-school average scores of the children in the TB instruction were all higher than the average score of the children in the school using SB instruction.

The school implementing the SB programme had implemented the TB programme prior to the switch to the SB programme that occurred 2 years earlier. Figure 5 displays the proportion of children reading at different score levels at the end of G1 for 3 cohorts of G1 children over the previous 4 years. The first and second sets of bars represent the scores of children in the first and second year cohorts when TB instruction was used in the school. The 3rd cohort is omitted. This was the year that the SB instruction was first implemented. The third



TB, text-based; SB, standards-based.

FIGURE 2: Comparison of mean scores of K children on Correct Letter Sounds, and Whole Words Read measures (*DIBELSnext*) by school, TB versus SB.

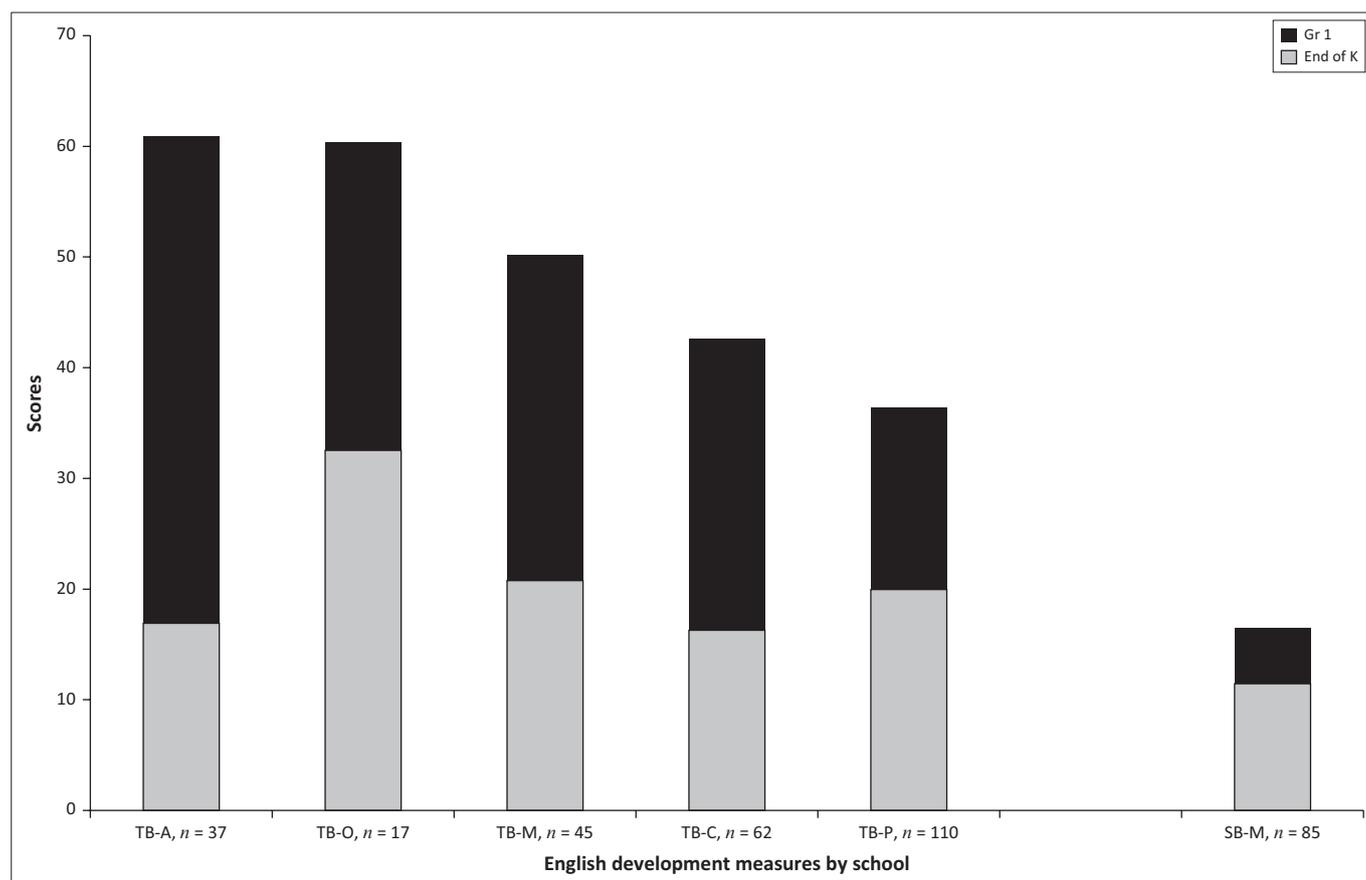


FIGURE 3: Change in mean scores of children by school on Correct Letter Sounds at end of K and end of G1 (only children with scores in both K and G1 are included).

TABLE 4: Comparison of mean scores on the Correct Letter Sounds measure at end of K and for the same children at the end of G1 for the TB and SB treatments. (Only children with scores in both K and G1 are included.)

Treatment group	n	K CLS (SD)	Gr1 CLS (SD)	Change (SD)
TB treatment	322	18.4 (14.2)	42.9 (37.7)	24.5 (30.6)
SB treatment	85	11.5 (11.1)	16.5 (18.6)	5.0 (14.4)
Pooled SD	-	13.9	36.2	29.1
Effect size	-	0.50	0.73	0.67
Probability	-	$t = 4.2, p < 0.0001$	$t = 6.3, p < 0.0001$	$t = 5.7, p < 0.0001$

TABLE 5: Comparison of mean scores of children learning from TB instruction versus SB instruction at end of G1 on DIBELSnext measures: Correct Letter Sounds (CLS), Words Correct Per Minute (WCPM) on the ORF and Percent Accuracy.

Treatment group	n	CLS (SD)	WCPM (SD)	% Accuracy (SD)
TB treatment	388	41.6 (37.1)	18.0 (23.4)	49% (31%)
SB treatment	105	17.5 (19.5)	7.0 (14.5)	28% (29%)
Effect size	-	0.87	0.60	0.70
Probability	-	$t = 6.4, p < 0.0001$	$t = 4.6, p < 0.0001$	$t = 6.2, p < 0.001$

set of bars (solid black) represents the performance of the 4th cohort, our study sample in the SB treatment.

During the period of the TB instruction in the school, approximately 30% of the children at the end of G1 could read only five words or less correctly. After implementing the SB model, 63% of the children could read only five English words or less correctly at the end of G1. Reading performance within the SB school had deteriorated dramatically with the implementation of the SB instructional model. The percentage of non-readers (children scoring in the range of 0 to 5 words correct) doubled at the end of G1 with the implementation of the SB model.

All 13 comparisons favoured the TB treatment and were well beyond the minimal test of statistical significance, $p < 0.004$ to correct for multiple comparisons using the Bonferroni method. ES were also quite large, indicating educational/practical significance.

Discussion and conclusion

The results of this study support the efficacy of TB structured immersion language as compared to the teacher-designed SB instructional model in all language and reading assessed. However, the lack of random assignment to treatment is a major weakness in the design of the study.

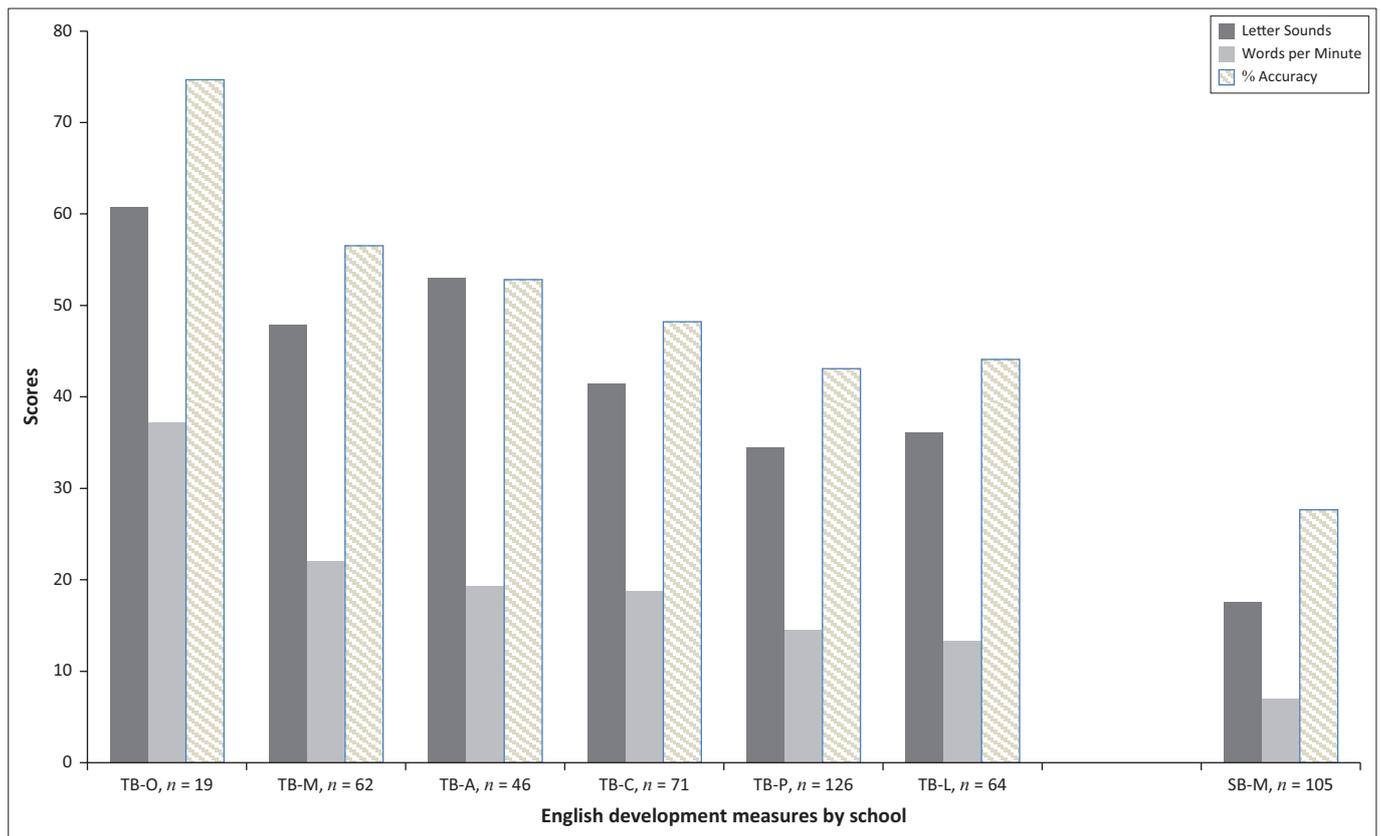


FIGURE 4: Average Correct Letter Sounds, ORF average Words Correct Per Minute and ORF average percent accuracy on end of year G1 *DIBELSnext* measures by school.

Several factors about the quasi-experimental design though, lend confidence to the findings. First, the population of American Samoa is very homogeneous. The children included in the sample for this study all began school speaking only very minimal English, if any at all. The *DIBELSnext* measure, commonly used to find variability in performance of K children on the American mainland, resulted in a universal score of zero in American Samoa. Prior to instruction, we did detect variability in the children's English competence through use of the very basic 'What's your name?' test. We found that any advantage of greater English competence at the start of K favoured the SB treatment. The results were in the opposite direction. This pattern of results would tend to rule out selection bias as a threat to internal validity. In addition, the large sample size, the consistency of the results across multiple comparisons, and the strength of the resulting differences between treatments all lend confidence to the findings of this study, in spite of the quasi-experimental design.

Assessment neutrality towards the treatments is difficult to prove. The fact that the English language measures were independent measures of the most basic expected outcomes, common to both treatments, rather than a test of final objectives, assures greater neutrality.¹ The CLS measure included the complete set of letters in English, which rules out bias toward one treatment or the other. Producing sounds for letters in the context of words or syllables, is a commonly

¹The content of the *Language for Learning* programme can be reviewed online. http://www.mheducation.com.sg/school/catalog/intl_gen_2010/Reading_Language-Arts-124-144

accepted emergent reading skill. *DIBELSnext* ORF measures had strong technical adequacy and were not aligned to any particular curriculum.

The key components embedded in the TB programme that explain its success are the following: The programme is highly structured, and prior knowledge of English is not assumed. The programme entails the introduction and practice of vocabulary in a carefully controlled manner, with clear expectations, focused tasks and direct explicit instruction. Ongoing progress monitoring is built into the programme. Receptive and expressive language proficiency is developed with great consistency across teachers and children due to the scripted, highly structured nature of the instruction.

The SB instruction depended on teachers to design their own lessons from descriptions of learning and performance indicators. The teacher presentations are expected to vary greatly from one teacher to the next. Certainly their presentations varied due to uncertainties the teachers had about the number of examples required to understand a concept, the amount of practice and the competence of the teacher to present accurate information.

Generalisability to the other contexts may be limited for several reasons. The children in the present study had no English competence at all when entering school. Indigenous children in other contexts may have more English competence, for example, Native American children and indigenous

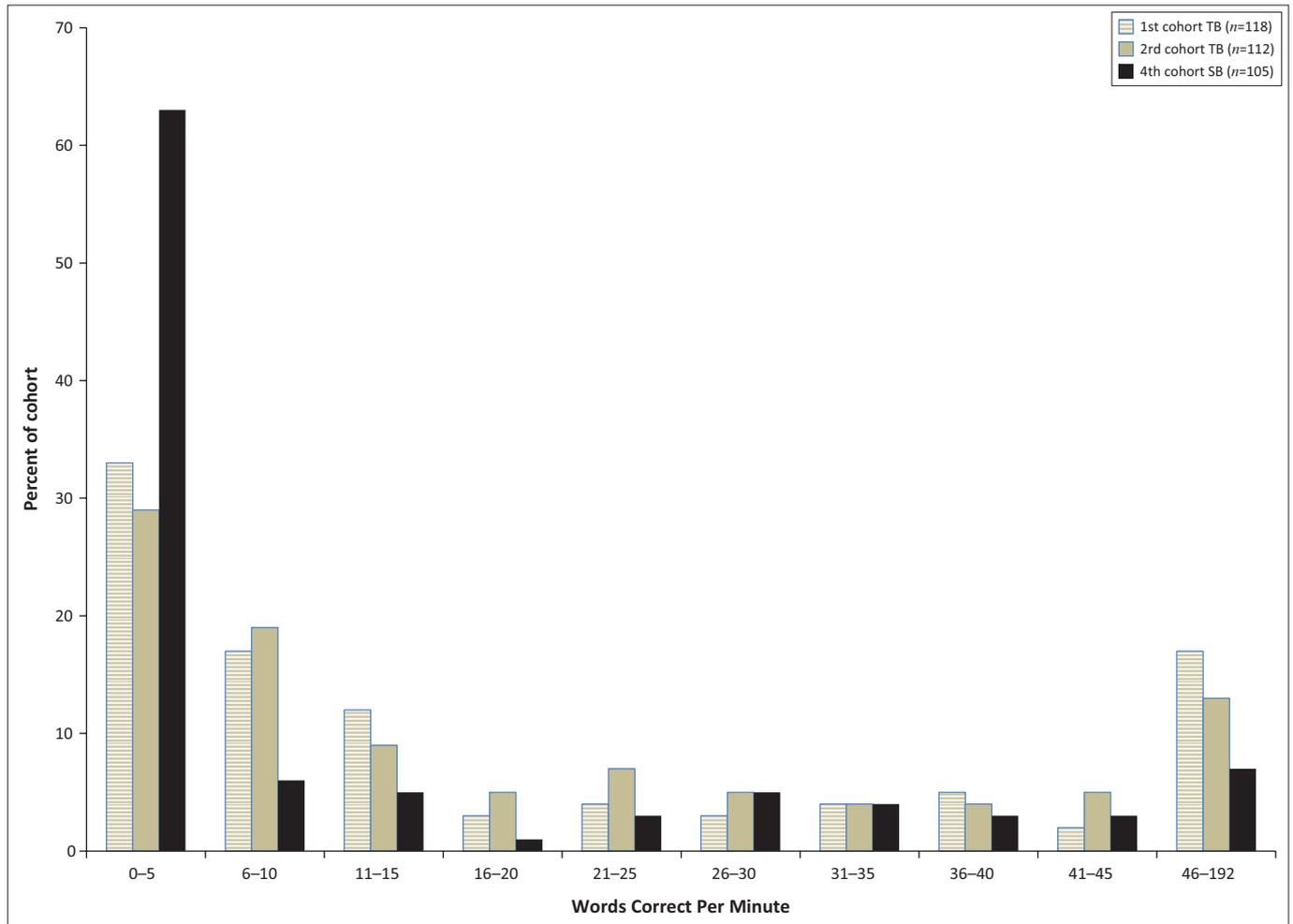


FIGURE 5: Comparison of the distribution of WCPM scores of 3 cohorts of G1 children attending the SB school. Cohorts 1 and 2 experienced TB instruction, Cohort 3 was omitted because it was the year of implementation of the SB treatment. Cohort 4 comprised the SB treatment group.

children in urban areas of South Africa or Australia. Children with more English competence may function better in a less structured environment for English language acquisition.

These indigenous teachers of indigenous children in American Samoa share characteristics of indigenous populations around the world, especially Australia, South Africa, the islands of the entire Pacific region, Native American populations and so on. The teachers' competence in English and lesson design is lacking. Many have not completed the standard teacher education programme but are hired because of an extreme shortage of well-trained teachers willing to work in these contexts. These indigenous teachers have much to offer in spite of less formal education. They understand the cultural context of the children, can speak their language and know how to motivate the children. With a high quality, proven TB programme, they can become much more effective. As the Grossen and Kelly (1992) study showed, the children taught by the indigenous teachers with the same proven TB programme outperformed the indigenous children taught by a native English speaker using a child-centred curriculum. This study replicates the results of the Grossen and Kelly (1992) study conducted in the context of South Africa. Effectiveness aside, the high level of control that the TB curriculum has over instruction introduces issues of power

and ethics. Many advocate child-centred approaches as the key to more democratic schooling. However, the exclusive use of child-centred approaches may fail to empower disadvantaged groups. We believe that 30 minutes of daily instruction that provides each and every child with the basic tool skills, as the TB instruction did, is more democratic than instruction that provides only the most able children with those tools, as the less-directive SB instruction did.

Acknowledgements

Competing interests

The authors declare that they have no financial or personal relationships which may have inappropriately influenced them in writing this article.

Authors' contributions

All three authors B.G., N.M. and J.N. contributed to the conceptualisation of the study. B.G. was the project leader, and she was responsible for data production and analysis. J.N. and N.M. made contributions to the data analysis process and the conceptualisation of the paper. The main tasks in writing the article, namely, the introduction and background, literature review, methodology, presentation of results and

discussion and conclusion was distributed equitably between the three authors. All three authors were involved in the proof-reading, editing and revision processes.

References

- Arizona Department of Education, 2004, *The effects of bilingual education programs and structured English immersion programs on student achievement: A large-scale comparison*, viewed 4 November 2014, from <http://epsl.asu.edu/epru/articles/EPRU-0408-66-OWI.pdf>
- Becker, W. & Gersten, R., 1982, 'A follow-up of follow through: The later effects of the direct instruction model on children in fifth and sixth grades', *American Educational Research Journal* 19, 75–92. <http://dx.doi.org/10.3102/00028312019001075>
- Brock-Utne, B., 2007, 'Language of instruction and student performance: New insights from research in Tanzania and South Africa', *International Review of Education* 53, 509–530. <http://dx.doi.org/10.1007/s11159-007-9065-9>
- Clark, K., 2009, 'The case for structured English immersion', *Educational Leadership*, 42–46.
- Cohen, J., 1969, *Statistical power analysis for the behavioral sciences*, Academic Press, New York.
- Cummins, J., 2009, 'Bilingual and immersion programs', in M.H. Long & C.J. Doughty (eds.), *Handbook of second language teaching*, pp. 161–181, Blackwell, London.
- Cummins, J. & Hornberger, N.H. (eds.), 2008, *Encyclopedia of language and education*, 2nd edn., vol. 5, Bilingual education, Springer Science/Business Media LLC, New York.
- Department of Education, 2008, *Foundations for learning*, Department of Education, Pretoria.
- Department of Education, 2011a, *Curriculum and assessment policy statement. Grades 1–3 – English second additional language*, Department of Education, Pretoria.
- Department of Education, 2011b, *Annual National Assessments (ANA). Grades 1–3 – English second additional language*, Department of Education, Pretoria.
- Dunn, O.J., 1961, 'Multiple comparisons among means', *Journal of the American Statistical Association* 56, 54–64. <http://dx.doi.org/10.1080/01621459.1961.10482090>
- Engelmann, S. & Osborn, J., 2008, *Language for learning*, McGraw-Hill Education, Columbus, OH.
- Faltis, K., 2011, *Bilingual, ESL, and English immersion: Educational models for limited English proficient students in Texas*, viewed 2 November 2014, from <https://publicpolicy.pepperdine.edu/policy-review/2011v4/content/bilingual-esl-english-immersion.pdf>
- Genesee, F., 1985, 'Second language learning through immersion: A review of U.S. programs', *Review of Educational Research* 55(4), 541–561. <http://dx.doi.org/10.3102/00346543055004541>
- Gersten, R., 1985, 'Structured immersion for language minority students results of a longitudinal evaluation', *Educational Evaluation and Policy Analysis* 7, 187–196. <http://dx.doi.org/10.3102/01623737007003187>
- Gersten, R. & Woodward, J., 1985, 'A case for structured immersion', *Educational Leadership*, September, pp. 75–79.
- Gersten, R. & Woodward, J., 1995, 'A longitudinal study of transitional and immersion bilingual education programs in one district', *Elementary School Journal* 95(3), 223–239. <http://dx.doi.org/10.1086/461800>
- Good, R.H. & Kaminski, R.A. (eds.), 2011, *Dynamic indicators of basic early literacy skills (DIBELSnext)*, 6th edn., Voyager Sopris Learning, Dallas, TX, viewed 24 October 2009, from <http://www.voyagersopris.com/assessment/literacy-assessment/dibels-next>
- Good, R.H., Kaminski, R.A., Dewey, E.N., Wallin, J., Powell-Smith, K.A. & Latimer, R.J., 2011, *DIBELS Next Technical Manual (Draft)*, Dynamic Measurement Group, Eugene, OR, viewed 13 January 2015, from <http://www.oregonrti.org/wp-content/uploads/2013/09/DIBELS-Next-Technical-Manual.pdf>
- Grossen, B. & Kelly, B.F., 1992, 'Using direct instruction to improve the effectiveness of teachers in South Africa', *South African Journal of Education* 12(2), 143–148.
- Hakuta, K., Butler, Y. & Witt, D., 2000, *How long does it take English learners to attain proficiency?* Discussion Paper Policy Report No. 2000-1, University of California Linguistic Minority Research Institute, Berkeley, CA.
- Harcourt, 2007, *Trophies*, Harcourt, San Diego, CA.
- Hayes, J., Salazar, J.J. & Vuckovic, G., 2002, *Evaluation of the structured immersion program*, Los Angeles Unified School District, viewed 31 October 2014, from http://notebook.lausd.net/pls/ptl/docs/page/ca_lausd/fldr_organizations/fldr_plcy_res_dev/par_division_main/research_unit/publications/reports/rpt%20sei%20y3.pdf
- Huebner, T., 1986, 'Vernacular literacy, English as a language of wider communication, and language shift in American Samoa', *Journal of Multilingual and Multicultural Development* 7(5), 393–411. <http://dx.doi.org/10.1080/01434632.1986.9994255>
- Resnik, D.B., 2005, 'Eliminating the daily life risks standard from the definition of minimal risk', *Journal of Medical Ethics* 31, 35–38. <http://dx.doi.org/10.1136/jme.2004.010470>
- Rossell, C., 2005, 'Teaching English through English', *Educational Leadership*, 32–36.
- Rossell, C.H., 2003, 'The near end of bilingual education', *Education Next* 3(4), 44–52.
- Sievert, J., 2007, *Evaluation of structured English immersion and bilingual education on the reading skills of limited English proficient students in California and Texas*, viewed 4 November 2014, from <https://digital.library.txstate.edu/bitstream/handle/10877/3655/fulltext.pdf>
- Taylor, S. & Coetzee, M., 2013, *Estimating the impact of language of instruction in South African primary schools: A fixed effects approach*, viewed 12 October 2014, from www.essa2013.org.za/fullpaper/essa2013_2586.pdf
- Voyager Sopris Learning, 2004, *Read well K to 3*, Voyager Sopris Learning, Dallas, TX.
- Zimmerman, I.L., Steiner, V.G. & Pond, R.E., 2011, *Preschool Language Scale – 5th edition., (PLS-5)*, Pearson, San Antonio, TX.