Bilingual education, metalinguistic awareness, and the understanding of an unknown language*

HAGAR TER KUILE
MICHIEL VELDHUIS
SUZANNE C. VAN VEEN
JELTE M. WICHERTS
University of Amsterdam

An increasing number of schools offer bilingual programs, where lessons are taught in more than one language. Several theories state that bilinguals have greater metalinguistic awareness than monolinguals. We investigated whether this greater metalinguistic awareness is also related to an increased ability to understand an unknown language. To measure metalinguistic awareness and the ability to understand text written in an unknown language, we designed the Indonesian Language Test (ILT). The ILT consists of items regarding a story in Indonesian. Dutch high school students from monolingual and bilingual classes were administered the ILT, a Dutch Language Test, an English Language Test, and a general intelligence test. The ILT showed promising psychometric properties. Bilingual students scored significantly higher on the ILT than monolingual students. Multi-group confirmatory factor analyses showed (i) that ILT measures the ability to understand an unknown language, and (ii) that bilingual students score significantly higher than monolingual students on this ability. Both observations support the notion that bilingual education increases metalinguistic awareness and therefore the ability to understand an unknown language.

Keywords: bilingual education, metalinguistic awareness, language acquisition, Indonesian Language Test

Introduction

An increasing number of secondary schools in the Netherlands offer bilingual programs, where lessons are taught in more than one language (Edelenbos & de Jong, 2004). Concerns have been raised about possible detrimental effects of bilingual education on the first language or on overall academic achievement (Lazaruk, 2007). However, several studies suggest that children who master two languages have better cognitive development, are better able to form concepts, are more flexible in their thinking, and have better control over their attention than children who master only one language (Bialystok, McBride-Chang & Luk, 2005). Several studies have found that bilingual children have better metalinguistic awareness than monolingual children (Ransdell, Barbier & Niit, 2006; Whitehurst & Lonigan, 1998). Metalinguistic awareness refers to the understanding that language is a system of communication, bound to rules, and forms the basis for the ability to discuss different ways to use language. The acquisition of metalinguistic awareness is the last stage

of language development of children. After this stage, children are able to think and talk about language itself (Whitehurst & Lonigan, 1998). The goal of this research is to study whether increased metalinguistic awareness in bilingually taught children is related to the ability to understand the writings in an unknown language.

Metalinguistic awareness allows reasoning and application of logic with language. For example, a person who has reached this stage is able to reason that a word appearing often in a story, and that always starts with a capital letter, is probably the name of the main character of the story. Metalinguistic awareness is related to a greater ability to discover connotations from paralinguistic clues, and to understand ambiguities in language (Edwards & Kirkpatrick, 1999). It is to be expected that this kind of reasoning with language facilitates the understanding of texts written in an unknown language. Klein (1995) found that multilinguals who learned English as their third or fourth language, learned the language faster than bilinguals who learned English only as a second language. Learning a third language was also found to have a positive effect on proficiency in a second language that was learned previously (Griessler, 2001). Thomas (1988) concluded that this advantage in learning a new language was due to better metalinguistic awareness in bilinguals as compared to monolinguals. The hypothesis in this study

Address for correspondence:
Jelte M. Wicherts, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB, Amsterdam, The Netherlands
j.m.wicherts@uva.nl

is therefore that the more languages one is fluent in, the better a person's metalinguistic awareness, and therefore the greater his or her understanding of text written in an unknown language.

In order to test this hypothesis, we designed a short test to measure metalinguistic awareness and the ability to understand an unknown language. The psychometric properties of this test were the topic of a pilot study (ter Kuile, Veldhuis & van Veen, 2006), which showed promising results. The Indonesian Language Test (ILT) is meant for participants who speak Indo-European languages (such as English and Dutch), but who have no knowledge of Indonesian. Indonesian was chosen as the test language because it is from another language family, ensuring that understanding of the text would not be due mainly to common words. The text is written in the Latin alphabet.

The ILT consists of a short story in Indonesian, with enough words translated in order to give a general idea of the story. Questions on content, grammar and structure are asked pertaining to the story. Metalinguistic awareness is hypothesized to be necessary in order to answer these questions. Readers must use the translated words and logically analyze the sentences. If the ILT indeed measures metalinguistic awareness, bilinguals should perform better on the ILT than monolinguals. The metalinguistic task was designed as a story for several reasons. Metalinguistic awareness has been found to be closely linked to reading comprehension (Zipke, 2007). Edwards and Kirkpatrick (1999) argue that metalinguistic awareness should be measured in a natural setting in which language is comprehended in a normal manner. Reading a story fulfills these criteria. Stories have also often been found to be a natural medium for language acquisition, and may increase test-taking motivation among high school students.

This study has two predictions. First of all, we expected bilinguals to have better metalinguistic awareness and thus to outperform monolinguals on the ILT, even when controlling for general intelligence. In addition, we expected the scores on the ILT to be related to the scores on other tests of language skills.

Method

Participants

The study was conducted at five high schools in the western part of the Netherlands. These high schools include both a bilingual program and a regular educational program in Dutch only. In the Netherlands there are by now 60 bilingual high schools that teach approximately 50% of the subjects in English and the other half in Dutch. Subjects that are often taught in English are, besides English itself, Geography, History and Biology (Edelenbos & De Jong, 2004). The comparison of pupils of these bilingual programs to monolingual

students provides an excellent quasi-experimental set-up to determine whether bilingual adolescents have an advantage in understanding a new language. The duration of the bilingual program is three years, and represents the first half of the pupils' high school career. The schools were selected through the website of the National Association for Bilingual Education in The Netherlands. Each high school was located in a different city. Data were collected in December 2006 and January 2007. Neither the schools nor the students were offered any compensation.

We employed a matrix format that consisted of four different types of classes in the third year of high school, which is equivalent to 9th grade in an American high school. This age group was chosen because in younger children the extent of development of metalinguistic awareness becomes a factor, while in older subjects factors not related to languages, such as differences in general knowledge, may confound the outcome. The Dutch high school system consists of several types, each having a different level of difficulty and requirements. All participants were pupils in the two highest levels. The highest levels are denoted collectively as "VWO", and are the only levels that allow direct access to the university. The highest level is called the Gymnasium, and its curriculum includes Latin and classical Greek. The second highest level is the Atheneum, which covers the same curriculum as the Gymnasium, with the exception of the classical languages. The other two foreign languages that are taught to all Gymnasium and Atheneum students are French and German. Classes that participated in our study were from these two different levels of education, bilingual as well as monolingual.

The four main conditions consisted of: (i) bilingual third-year Gymnasium classes (Latin and Greek, and 50% of the classes are taught in English), (ii) bilingual third-year Atheneum classes (50% of the classes taught in English), (iii) the monolingual third-year Gymnasium classes (Latin and Greek, classes taught in Dutch only), and (iv) the monolingual third-year Atheneum classes (classes are taught in Dutch only). Bilingual Gymnasium students receive the most language education, and are therefore predicted to have the highest scores on the ILT, followed by students in the bilingual Atheneum, monolingual Gymnasium, and monolingual Atheneum, respectively.

The 304 participants who were included in the analyses ranged in age from 12 to 16 years ($M = 14.30$, $SD = 0.57$). There were 130 males and 153 females (21 students did not indicate their gender). These participants were divided over four conditions, as shown in Table 1. There were no significant differences in age or sex between the conditions.

Several additional participants were excluded from the analyses. The data from one class of 27 students were unusable because of non-compliance with the standardized time limit. Three students who were fluent

Table 1. *Repartition over the conditions.*

Condition	N	Mean age	% of females
Atheneum	113	14.37	55.9
Gymnasium	39	14.24	52.0
Bilingual Atheneum	91	14.20	62.2
Bilingual Gymnasium	61	14.24	48.1

in Indonesian were excluded from the analysis. Eleven additional students were excluded because they had at least 10 missing answers on the ILT, or because they had provided only one correct or non-missing answer to the 15 questions in the Indonesian Language Test, which indicated that they had not participated seriously. Finally, data from 19 students with a dyslectic disorder were excluded, as this may have influenced their test scores on language skills.

Materials

Four tests were used in this study: the English Language Usage Test (ELT), the Dutch Language Usage Test (DLT), Raven's Advanced Progressive Matrices Test, the Indonesian Language Test, and an exit interview.

The ENGLISH LANGUAGE TEST that was used in this study was a shortened version of the "Language Usage" subtest of the Differential Aptitude Test (DAT; Bennett, Seashore & Wesman, 1947). The questions used in the shortened version were chosen in such a way as to ensure that an as wide as possible range of grammatical rules were tested. This shortened test consisted of 15 sentences, each divided into four parts lettered A, B, C and D. Most sentences contain a grammatical error, and the participants had to find this error and circle the corresponding letter on their answer sheet. For example:

Bill done / everything that / the boss asked / him to.
    A   B   C   D

If there was no error in a sentence, the participants were asked to circle the N on their answer sheet. The participants received one point for every correct answer, and no points for every incorrect answer. The maximum total score was 15 points and the minimum score was 0 points.

We used a shortened version of the DUTCH LANGUAGE TEST based on the "Zinnen" subtest of the Dutch version of the DAT (DAT-83; Evers & Lucassen, 1983). The item–total correlations of the original test from the standardization sample were used to determine which questions to include in the shortened version of 15 items. The format of this test is identical to the English Language Test. The maximum total score of the DLT was 15 points and the minimum score was 0 points.

The INDONESIAN LANGUAGE TEST (ILT; ter Kuile, Veldhuis & van Veen, 2006) is designed to measure the ability to understand an unknown language. The test consists of a story of 18 lines and 180 words written in Indonesian. The story was written especially for this test by K. S. Sumbayak, an English teacher in Jakarta. In Indonesia it would be suitable for children aged 6–9 years. The story is about a boy who chooses to go to a soccer match instead of doing his homework assignment. In the text there are 28 words written in bold. The Dutch translations of these words are given on the next page. The words were chosen in such a way as to give a general idea of what the story is about, but the remaining text must be understood using general logic and metalinguistic skills. Participants have to answer 15 questions about specific parts of the text. There are three different types of questions: content, grammar, and structural questions. The ILT is included in the Appendix. Each question is worth a different number of points, varying from 0 to 2 points. The minimum total score is 0 and the maximum is $20\frac{1}{2}$. Each test was scored by three independent raters. The participant's final ILT score is the average of the scores from the three raters.

General intelligence was measured by a shortened version of Set II from Raven's ADVANCED PROGRESSIVE MATRICES (Raven, Raven & Court, 1998). The Raven's test was chosen to measure general intelligence for several reasons. First, the Raven's has been found to correlate highly with the general intelligence factor (Carroll, 1993; Jensen, 1998) yet it can be administered in a relatively short time. Second, this test relies heavily on the ability to use logical reasoning. This ability may also be used to answer the questions on the ILT. Although we did not administer a full IQ battery, the Raven's Test can be used as a reasonable control for the effect of general reasoning ability. The complete test consists of 36 items, the first 20 of which were used in the current study. The problems in this test consist of a matrix of nine figures with a logical pattern from left to right and from top to bottom, and one figure missing in the bottom right corner. The participants have to choose from eight alternative figures which figure will logically complete the pattern. The items are presented in increasingly difficult order. Participants were given a time limit of 10 minutes to complete the shortened Raven's Test. Participants received one point for every correct answer, and no points for incorrect answers. Thus, the scores on this test varied from 0 to 20 points.

The EXIT INTERVIEW consisted of questions about pupils' age, sex, their nationality and that of their parents, which languages they spoke fluently, whether they were dyslectic, and whether they knew any Indonesian.

Reliability of the tests
Cronbach's Alpha reliabilities were as follows: Dutch Language Test: .66, English Language Test: .55, Raven's

Test: .81, and the ILT: .70. The inter-rater reliabilities of the ILT scores were .92, .93, and .95. All reliabilities were considered sufficiently high for the present purposes, although the ILT may be lengthened in the future to improve reliability.

Procedure

The tests were group-administered in a regular classroom. The participants were first given a brief oral introduction in Dutch by one of the three testers. The participants were told that they would be making several language and cognitive skills tests. They would not be receiving a grade for these tests, but they were requested to cooperate in a serious and motivated manner, and to work individually. The tests were completed anonymously, but the test booklets were numbered. The students were told they could use this number later to ask for their test results or have their data removed from the analysis. Students found an answer booklet and a battery of tests on their table, sorted in the order in which the tests were taken; the Dutch Language Usage Test, the English Language Usage Test, the Raven's Test, the Indonesian Language Test, and an exit interview. Before starting each test, a tester would read out loud the instructions that were written on the first page of every test, and tell participants what the time limit was. After that they were allowed to begin. The time limits were 5 minutes for the DLT, 5 minutes for the ELT, 10 minutes for the Raven's Test, and 15 minutes for the ILT. All together administering the tests took about 40 minutes. Two minutes before the end of every time limit the participants received a warning from the tester. After completion of the tests, students were thanked for their cooperation and asked to fill in the exit interview. The administration of the tests was the same for all classes.

Statistical analyses

We computed reliabilities, correlations, and used MANCOVA on the scaled scores, with Raven's Test scores as a covariate to correct for general intelligence. In addition, we used confirmatory factor analysis to test for the dimensionality of the ILT. We employ multi-group confirmatory factor analysis with mean structure to test whether bilinguals and monolinguals differ in the ability to understand an unknown language or only in the specific languages. To this end, we test alternative factor models in which the differences between monolingual and bilingual groups on different languages are either due to the common factor ability to understand an unknown language and/or to skills in the three specific languages.

Results

A MANOVA was conducted with the scores on the ELT, the DLT and the ILT as dependent variables, the scores

Table 2. *Means and standard deviations of all tests per condition.*

Test	Condition	M	SD	N
Indonesian	Bilingual Gymnasium	9.75	3.30	54
Language Test	Bilingual Atheneum	8.21	3.24	79
	Gymnasium	8.06	3.69	46
	Atheneum	7.33	3.13	125
Dutch	Bilingual Gymnasium	7.93	2.89	54
Language Test	Bilingual Atheneum	6.90	2.81	79
	Gymnasium	7.05	2.64	41
	Atheneum	6.53	2.53	120
	Bilingual Gymnasium	7.96	1.92	54
	Bilingual Atheneum	7.08	1.96	79
	Gymnasium	5.76	1.82	46
	Atheneum	5.62	2.16	122
Raven's Test	Bilingual Gymnasium	15.98	2.98	54
	Bilingual Atheneum	14.18	3.88	79
	Gymnasium	15.26	3.04	46
	Atheneum	14.18	4.07	125

on the Raven's Test as covariate, and as between-subjects factors; mono- or bilingual education and Atheneum or Gymnasium education. Results of this MANOVA (after list-wise deletion of missing data $N = 294$) showed that after correction for the differences in general intelligence test performance, the scores of the students in the bilingual classes were significantly higher than the scores of the students in the monolingual classes on the Indonesian Language Test ($F(1,289) = 10.76$, $p = .001$). This supports the hypothesis that bilingual students are better at understanding an unknown language. Means per group for all tests are reported in Table 2. Due to missing data, the Ns differ per test.

The classes scored in the order that was predicted according to the amount of language education included in the curriculum. This was from highest to lowest: bilingual Gymnasium (M = 9.75), bilingual Atheneum (M = 8.21), Gymnasium (M = 8.06), and Atheneum (M = 7.33). The main effect for Atheneum vs. Gymnasium on the Indonesian Language Test was marginally significant: $F(1,289) = 2.88$, $p = .091$.

As was expected, the bilingual students also scored significantly higher on the English Language Test ($F(1, 289) = 46.91$, $p < .001$). There was no significant main effect of mono-or bilingual classes on the Dutch Language Test ($F(1,289) = 2.61$, $p = .107$). This was consistent with our hypothesis.

Dimensionality of the ILT

We studied the dimensionality of the 15 items of the ILT by fitting a one-factor confirmatory factor analysis

Table 3. *Correlations between the different measures.*

	ELT	DLT	ILT	Raven's Test
English Language Test	1	.285	.206	.140
Dutch Language Test	.424	1	.249	.206
Indonesian Language Test	.352	.338	1	.214
Raven's Progressive Matrices Test	.257	.313	.301	1

Note: All correlations significant (p < .01, two-tailed). Correlations below diagonal for all participants combined ($N = 294$); Correlations above diagonal are pooled within-class correlations (13 classes).

Figure 1. Confirmatory factor model of the factor ability to understand an unkown language.

model on the data pooled across classes. Because of non-normality of items scores, we employed maximum likelihood estimation with Satorra-Bentler corrected χ^2. In terms of Hu and Bentler's (1999) cut-off criteria, the one factor model fitted well ($N = 304$): $\chi^2(DF = 90) = 143.5$, $p < .001$, RMSEA = .044, CFI = .93, SRMR = .054. Standardized factor loadings for the ILT items are given in the Appendix.

Multi-Group Confirmatory Factor Analyses

The correlations between the four tests are shown in Table 3. Below the diagonal are the correlations of all participants, above the diagonal we display the pooled within-class correlations. The correlations for all participants are significant at the p < .01 level and lie between .25 and .45, thereby indicating a moderate relation between scores on all tests. The pooled within-class correlations are significant as well, but turn out to be slightly lower (.14–.29). The confirmatory factor model is given in Figure 1. This model was tested in multi-group analyses, in which each class represented one group. Hence there are 13 groups in this analysis. General intelligence acted as a covariate in the model by restricting the residual variance of the Raven's Test at zero. We note that because of the unidimensional nature of the Raven's Test, the general intelligence factor cannot be modeled as a latent construct. However, because this test has been found to correlate highly with the general intelligence factor (Carroll, 1993) we consider this approach to be a reasonable approximation. We fitted this model with and without across group restrictions. To study the nature of between-class differences in scores on the three language tests, we tested for measurement invariance across groups (Meredith, 1993), whereby measurement invariance would mean (1) that the relation between language test scores and the underlying factor (i.e., ability to understand an unknown language) is identical across the different (types of) classes, and (2) that between-
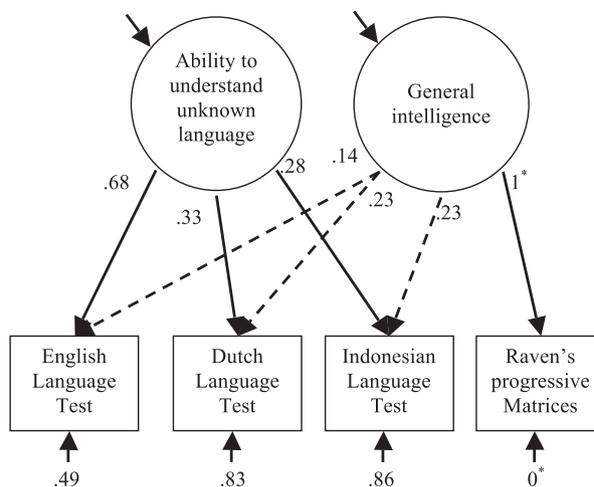
class differences could be explained in terms of group differences in the underlying ability rather than to the unique language abilities tapped by the specific language tests (see Lubke, Dolan, Kelderman & Mellenbergh, 2003; Wicherts, Dolan & Hessen, 2005). Measurement invariance was tested by restricting, in a stepwise manner, factor loadings, residual variances, and intercepts to be invariant across groups. If, for instance, the bilingual classes lag behind in Dutch language proficiency, then we would expect that the intercept associated with the Dutch Language Test would not be invariant across the bilingual and monolingual classes, resulting in a violation of measurement invariance. On the other hand, if the different types of classes differ merely in the ability to understand an unknown language, we would expect measurement invariance across those types of classes in the model depicted in Figure 1. Thus, the measurement invariance analyses can be used to shed light on the sources of the difference between monolingual and bilingual students in the different language abilities. As articulated by Meredith (1993), possible differences in factor variances can be due to differences between classes in the variance of the latent factor (i.e., differences in factor variances are not a matter of MEASUREMENT invariance) and therefore factor variances are freely estimated across classes throughout the multi-group analyses.

Figure 1 includes the mean of the standardized parameter estimates across the classes in the most-restricted model. As can be seen, the ILT taps both general intelligence and the ability to understand an unknown language, with the most variance explained by the latter factor. The fit measures in Table 4 indicate that the restrictions of invariance did not result in deteriorations of model fit (all $\Delta\chi^2$s: $p > .10$). This suggests that there is measurement invariance across

Table 4. *Fit measures of multi-group factor analyses.*

Restriction	DF	χ^2	p	RMSEA	AIC
Factor loadings	60	58.8	.518	.000	300
Residual variances	96	96.6	.463	.000	263
Intercepts	120	124.1	.381	.000	240
Means of language factor zero in all classes	132	226.3	.000	.136	285
Means of language factor zero in bilingual classes	126	183.3	.001	.091	260
Means of language factor zero in monolingual classes	126	169.1	.006	.084	257
Means of language factor zero in monolingual classes except one outlying class	125	145.1	.105	.051	246

DF = Degrees of Freedom; RMSEA = Root Mean Square Error of Approximation; AIC = Akaike Information Criterion
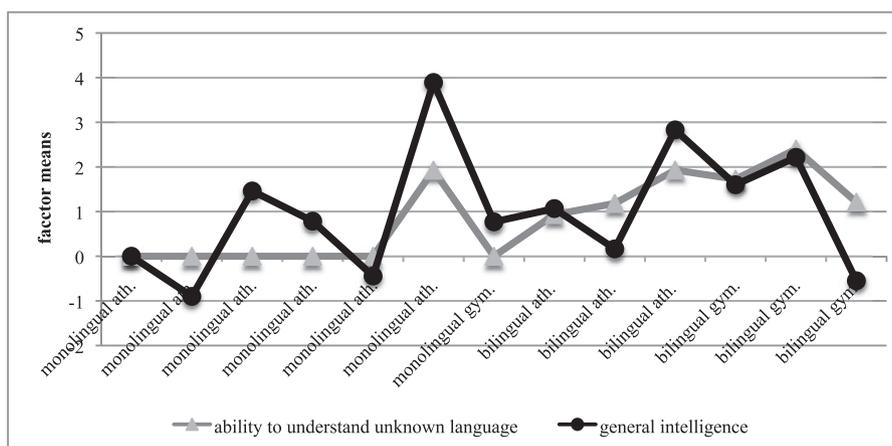

Figure 2. Factor means of the classes in measurement invariant multi-group confirmatory factor model.

these classes and that differences in mean scores on the three indicators of the ability to learn an unknown language between the classes can be safely interpreted in terms of group differences in the latent factor metalinguistic awareness (and general intelligence). After we established measurement invariance across groups, we restricted the factor means of the ability to understand an unknown language to zero across all classes. This resulted in a severe deterioration in model fit: $\Delta\chi^2$ ($DF = 12$) = 102.2, $p < .001$. Restricting the factor means to zero in the bilingual classes only resulted in a worsening of model fit as well: $\Delta\chi^2$ ($DF = 6$) = 59.2, $p < .001$. The same restriction in the monolingual classes ($\Delta\chi^2$ ($DF = 6$) = 45.0, $p < .001$) also resulted in a non-fitting model in terms of RMSEA and absolute fit. However, we found that the misfit was almost entirely due to one high-scoring monolingual Atheneum class. After we freed this parameter, the model fitted well in terms of RMSEA and in terms of overall exact fit. The means of the general intelligence and the ability to learn an unknown language are displayed in Figure 2. The factor mean differences can be interpreted in terms of the within-class factor

variance, which is 1.12 on average across the 13 classes (i.e., the within-class SD of the factor scores is around 1.06). As can be seen, the language skills factor mean of all the bilingual classes were higher than those in the monolingual classes (bar the outlying class). Wald tests showed that the language skills factor means of these bilingual classes were all significantly higher than the corresponding factor mean (restricted at zero) in the monolingual classes; all Zs > 2.79, $p < .01$. Thus, these analyses suggest that differences between the monolingual and bilingual classes are due to differences between the classes in the ability to learn an unknown language, rather than to ability in specific languages.

Discussion

Our results suggest that bilinguals have a better ability compared to monolinguals to understand an unknown language, arguably due to their greater metalinguistic awareness. Bilingual education therefore seems to give students an additional advantage above simply being fluent in two languages. This conclusion supports

previous findings that bilinguals have better metalinguistic awareness (Campbell & Sais, 1995; Ransdell, Barbier & Niit, 2006). In addition, the data imply that this ability improves language acquisition. Students who spoke more languages at home also showed this increased ability. The classes scored on the ILT in the order of proficiency and the number of languages learned, as we predicted.

We are therefore confident that the Indonesian Language Test (ILT) measures the ability to understand an unknown language, and does not measure the language skill in any one specific language, also because only students who scored high on both the Dutch AND English Language Tests scored high on the ILT. Raven's Test performance also correlated positively with the ILT, as was expected, but the bilingual students still scored higher than the monolingual students when the ILT scores were corrected for the scores on the cognitive ability test. It can therefore be ruled out that cognitive ability (as measured by the Raven's Test) is the sole reason for higher scores on the ILT.

To further confirm that a high score on the ILT requires metalinguistic awareness, future research can investigate the correlations between the ILT and specific measures of metalinguistic awareness that do not use an unknown language, such as the measures used by Campbell and Sais (1995) or Zipke (2007).

The difference between the Gymnasium classes and Atheneum classes was only marginally significant. This is probably due to the fact that there were two Atheneum classes who had relatively high average scores on the ILT. However, these classes also had high scores on both the English and Dutch Language Tests. That is to say, the students in these classes were not in the bilingual program but they could in fact be considered to be bilingual.

It is possible, and even probable, that some pre-selection takes place when students choose to study in a bilingual program. These students may already have greater metalinguistic awareness than students who choose to go to a regular high school. However, the fact that those who were bilingual or multilingual through upbringing also showed an increased ability to understand an unknown language, and that Atheneum classes that were proficient in at least two languages also exhibited

this ability, means that the effect of pre-selection is not enough to explain the results.

The bilingual classes also scored higher on the Dutch Language Test than the monolingual classes. This replicates the findings from the study by Griessler (2001), who found that learning a third language facilitates the language skills in the second language. The metalinguistic awareness gained by learning a second language may also improve the language skills in the first language. In the early stages of bilingual education it was feared that learning another language would be detrimental for the mother language. This fear seems ungrounded.

Earlier studies, especially the study by Bialystok, McBride-Chang and Luk (2005), found a positive influence of bilingualism on the cognitive development of children. In this study, no difference in cognitive ability test performance was found between the bilingual and monolingual students. The focus of this study, however, was not on cognitive development. The test used to measure cognitive ability was as a result relatively short and mainly measured the ability to think logically. This was chosen because it is possible that logic can be used to help answer the questions on the ILT. Hence it is possible that differences in other domains of intelligence were not reflected in the scores on this particular test.

The Indonesian Language Test was found to be a reliable test that measures primarily one construct, viz. the ability to understand an unknown language. The scores on the ILT correlate highly with other language skills test scores. Such a test offers some interesting possibilities. Specifically, with the ILT it may be possible to predict whether someone will be able to easily learn new languages. Not only schools that offer bilingual education could very well use this test, but regular schools too, to assess for example whether children are able to participate in Gymnasium programs. Work sectors in which language ability is very useful, for example the Foreign Service or companies operating worldwide, could use the ILT to examine applicants. However, although the ILT shows promising psychometric properties, increasing the length of the story and the number of questions would increase the reliability and add to the predictive power of the ILT. Further research should therefore be conducted to help make the ILT more reliable and practical to use.

Appendix. Indonesian Language Test

Instruction

This test consists of an Indonesian text of 18 lines, a list of some of the words in the text with their translations and 15 questions on the text.

Try to answer these questions as good as you can, using the list of words with translations and your knowledge on other languages.

Understanding this text is not easy, but with the list of words with their translations it is feasible.

Take your time to think about the questions, but keep the time limit in mind.

You've got 15 minutes to complete this test.

Good luck!

HUKUMAN BUAT EDO

Edo, Miko, Rima, Lala, dan Dimas **berjanji** akan **belajar** bersama di rumah Lala. **Kakak** Lala yang kuliah di Bandung mau **membantu** mereka mengerjakan **tugas** Bahasa Indonesia.

Sepulang sekolah, Edo, Rima, dan Lala **pulang** bersama karena rumah mereka satu (1)
5   arah. di tengah perjalanan, Edo melihat **pengumuman**.

"**Sore** ini ada pertandingan **sepak bola** di lapangan kelurahan. Aku harus **menontonnya**!" kata Edo.

Lala **mengingatkan** bahwa sore itu mereka harus belajar bersama di rumahnya.
"Aku tidak ikut belajar bersama karena pertandingan ini **lebih menarik**!" kata Edo.

10   Sore harinya, Miko, Rima, Lala, dan Dimas belajar bersama **sedangkan** Edo sedang **bersorak-sorak** menonton pertandingan sepak bola. Hati Edo sangat **senang** karena **regu** idolanya menjadi **pemenang**. Sampai di **rumah**, Edo merasa lelah. Ketika jam menunjukkan jam delapan malam, Edo sudah **tertidur** pulas.

Esok harinya, Edo datang tepat pada saat **bel berbunyi**. Guru bahasa Indonesia
15   **masuk** ke kelas. Edo ingat kalau ia **belum mengerjakan** tugas. Edo pun mendapat hukuman. Edo harus **berdiri** di depan kelas **sampai** pelajaran Bahasa Indonesia **selesai**. Dalam hati Edo berjanji akan berdisiplin. Edo **tidak** akan pernah **lupa** lagi mengerjakan tugas sekolah.

© K. S. Sumbayak 2006

HUKUMAN BUAT EDO – PUNISHMENT FOR EDO

1	**berjanji**	–	promise
	belajar	–	study
2	**Kakak**	–	older brother/sister
	membantu	–	to help
	tugas	–	assignment
4	**Pulang**	–	to go home
5	**Pengumuman**	–	advertisement
6	**Sore**	–	afternoon
	sepak bola	–	soccer
7	**menontonnya**	–	to see
8	**mengingatkan**	–	(i) to remember
		–	(ii) to warn

9	**lebih menarik**	–	more interesting
10	**sedangkan**	–	while
11	**senang**	–	happy
	bersorak-sorak	–	cheer
12	**regu**	–	the team
	rumah	–	home
	pemenang	–	winner
13	**tertidur**	–	to sleep
14	**Esok harinya**	–	the next day
	bel berbunyi	–	the bell rung
15	**masuk**	–	to come in
	belum mengerjakan	–	hadn't done yet
16	**berdiri**	–	stand
	sampai	–	until
17	**selesai**	–	over (was)
	tidak	–	not
	lupa	–	to forget

Questions about the text "Punishment for Edo"

Item	Loading	Question
1.	.12	Who is the main character?
2.	.14	A lot of names are mentioned in the text, explain what kind of relationship these people have with each other.
3.	.24	At whose house did they want to study?
4.	.42	At a certain point Edo sees an advertisement. What was this advertisement for? Pay attention to the word *pertandingan* (line 6)
5.	–.07	In line 9, Edo reacts on Lala, find the subject and the verb in this sentence *aku tidak* (. . .) *belajar bersama*.
6.	.48	The word **senang** (r. 11) signifies *happy*, why is Edo happy?
7.	.13	The word *lelah* (r. 12) stands for a mood, which?
8.	.58	**Esok harinya** (r. 14) means *the next day*, it points to a turning point in the story. Where's Edo the next day?
9.	.61	In line 14 Edo remembers (*Ingat*) something, what does Edo remember? Find the word in the first paragraph that indicates this.
10.	.33	Give all the verbs from the sentence "*Sore harinya* (. . .) *sepak bola*" (lines 10, 11).
11.	.43	The sentence "*Hati Edo* (. . .) *menjadi pemenang*" (r. 11, 12) is made up of a principal and a subordinate clause. Which word connects the two parts of the sentence?
12.	.68	*Guru bahasa Indonesia* **masuk** *ke kelas*. (r. 14, 15) Translate this sentence.
13.	.60	The title of this text **hukuman buat edo** means **punishment for Edo**, what's Edo's punishment?
14.	.41	The last word of the text is *sekolah*, what is the meaning of this word?
15.	.79	Indonesian texts often have a strong moralising message, the writer wants to make some point. What do you think is the moral to this story?

Note: Factor loadings from 1-factor model (N = 304).

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1947). *Differential aptitude tests*. New York: Psychological Corporation.

Bialystok, E., McBride-Chang, C., & Luk, G. (2005). Bilingualism, language proficiency, and learning to read in two writing systems. *Journal of Educational Psychology*, *97*, 580–590.

Campbell, R., & Sais, E. (1995). Accelerated metalinguistic (phonological) awareness in bilingual children. *British Journal of Developmental Psychology*, *13*, 61–68.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Edelenbos, P., & de Jong, J. H. A. L. (2004). Vreemdetalenonderwijs in Nederland: Een situatieschets [Second language education in The Netherlands: An overview]. http://www.nabmvt.nl/publicaties/00013/ (retrieved January 12, 2007).

Edwards, H. T., & Kirkpatrick, A. G. (1999). Metalinguistic awareness in children: A developmental progression. *Journal of Psycholinguistic Research*, *28*, 313–329.

Evers, A., & Lucassen, W. (1983). *DAT'83 Differentiële Aanleg Testserie* [Differential Aptitude Test series]. Lisse, The Netherlands: Swets & Zeitlinger.

Griessler, M. (2001). The effects of third language learning on second language proficiency: An Austrian example. *International Journal of Bilingual Education and Bilingualism*, *4*, 50–59.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Klein, E. C. (1995). Second versus third language acquisition: Is there a difference? *Language Learning*, *45*, 419–465.

Lazaruk, W. (2007). Linguistic, academic, and cognitive benefits of French immersion. *The Canadian Modern Language Review*, *63*, 605–628.

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model. *Intelligence*, *31*, 543–566.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.

Ransdell, S., Barbier, M., & Niit, T. (2006). Metacognitions about language skill and working memory among monolingual and bilingual college students: When does multilingualism matter? *The International Journal of Bilingual Education and Bilingualism*, *9*, 728–741.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Advanced Progressive Matrices*. Oxford: Oxford Psychologists Press.

Ter Kuile, H., Veldhuis, M., & van Veen, S. C. (2006). De invloed van tweetalig onderwijs op het begrip van een onbekende taal [The influence of binlingual education on the understanding of an unknown language]. Internal report, Psychology Department, University of Amsterdam.

Thomas, J. (1988). The role played by metalinguistic awareness in second and third language learning. *Journal of Multilingual and Multicultural Development*, *9*, 235–246.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, *68*, 848–872.

Wicherts, J. M., Dolan, C. V., & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality and Social Psychology*, *89*, 696–716.

Zipke, M. (2007). The role of metalinguistic awareness in the reading comprehension of sixth and seventh graders. *Reading Psychology*, *28*, 375–396.