4. Results

4.1. Preliminary analyses

This section will provide the necessary background information for interpreting the analyses conducted within this study. First, a thorough investigation of the utilized instruments is presented. Factor analyses, measurement stability, and comprehensive scale correlations calculated separately for each group make up the section on instrument characteristics. The exploration of the instruments' psychometric properties and relationships across measures allows for a more comprehensive understanding of the instruments and constructs under investigation. The second section on sample descriptives gives a more detailed picture of the language abilities and learning behaviors of participants in both groups as well as a series of analyses aimed at illustrating the bilingual capacities of participants in the Turkish bilingual group. The analyses regarding the dual language abilities of the bilingual sample provides insight into what "bilingualism" actually entails within this sample.

4.1.1. Instrument characteristics

Since the verbal instruments used to collect information on phonological awareness, short-term verbal memory, vocabulary, and listening comprehension were modified to fit the specific needs of this sample and the particular constraints of this study, it is useful to consider the scale intercorrelations and validity in greater detail for the individual measures within each group. This section aims to uncover any discrepancies between the two groups in terms of scale constructs or functioning by testing for differential correlations among the scales. For all instruments possible, the factor analytic structure of the measures is also examined for both groups separately. In some cases, the analyses in this section also attempt to indicate the external validity of the measures. It was essential that the instruments used for this study functioned with equal precision and viability for both the Turkish bilingual and the German monolingual children.

Phonological awareness

Since the instrument measuring phonological awareness was modified considerably from its original form, a thorough process of evaluation was undertaken to examine the construction and inter-relations of the four subscales: pseudoword segmentation, vowel replacement, word remainder determination, and sound categorization. In the first set of analyses, concurrent correlations were calculated for each group at each time of measurement with Pearson product-moment correlation coefficients. This provided insight into the extent to which the scales correlated with each other and if these correlations differed between the monolingual and bilingual groups.

Table 12

Concurrent Correlations among Phonological Awareness Subscales at T1 and T2 for the Turkish Bilingual (TB) and German Monolingual (GM) Groups (Correlations at T1 shown below the diagonal; correlations at T2 shown above the diagonal)

	Pseudoword segmentation		Vowel replacement		Word remainder determination		Sound categorization	
	TB	GM	TB	GM	TB	GM	TB	GM
Pseudoword segmentation	--	--	.32**	.50**	.23*	.30*	.32**	.33**
Vowel replacement	.47**	.62**	--	--	.35**	.37**	.42**	.25*
Word remainder determination	.53**	.48**	.44**	.55**	--	--	.38**	.20
Sound categorization	.28**	.46**	.45**	.36**	.42**	.57**	--	--

Note. Using a Bonferroni approach to control for Type 1 errors across the 8 correlations results in a *p*-value cutoff of .006 to indicate a significant correlation.
*p < .05. **p < .006 (Bonferroni cutoff).

In Table 12, the concurrent correlations are provided for both the Turkish bilingual and German monolingual groups. For the German group, the phonological awareness scales correlated highly with one another, from $r = .36$ to $r = .62$ at T1. Slightly weaker correlations were found between the scales at T2 ($r = .20$ to $r = .50$) in the German group. Although somewhat lower than in the German monolingual group, all scales correlated significantly with each other at each time of measurement for the Turkish bilingual group as well (correlation coefficients ranged from .28 to .53 at T1 and from .23 to .42 at T2).

The Fisher *r* to *z* transformation formula was used to test for differences in correlation coefficients between the two groups. Although the correlation coefficients appear somewhat lower in the Turkish group in several cases, no significant differences emerged. It can therefore be assumed that the phonological awareness scales inter-correlate with one another in similar ways for Turkish bilingual and German monolingual second graders in this sample.

The next set of analyses addresses the issue of stability across time and longitudinal correlations among the four phonological awareness scales. Again, correlation coefficients have been calculated for each group separately to test for group differences in scale inter-correlations for the phonological awareness instrument. The six-month stability correlation coefficients are displayed in bold on the diagonal for both groups in Table 13.

For the German monolingual group, all scales showed relatively high stability as they significantly correlated with themselves six months later with correlations ranging from $r = .34$ to $r = .72$. Observed over a six-month period, the scales correlated moderately to highly with one another from T1 to T2. Of the 16 correlations within the German monolingual group, 13 were significant at a $p < .05$ level. These correlations can be interpreted as an indication of the general stability of the phonological awareness scales.

A similar pattern emerged within the Turkish bilingual group. Although the pseudoword segmentation scale exhibited somewhat lower six-month stability than the other scales, all scales seem to show sufficient stability ($p < .05$) over time with stability correlations ranging from $r = .21$ to $r = .58$. It is also useful to note that in the Turkish bilingual group, all four scales significantly correlated with one another from the middle of second grade (T1) to the end of second grade (T2).

Table 13

Longitudinal Correlations among Phonological Awareness Subscales at T1 and T2 for the and Turkish Bilingual (TB) and German Monolingual (GM) Groups

	Subscales at T2							
	Pseudoword segmentation		Vowel replacement		Word remainder determination		Sound categorization	
Subscales at T1	TB	GM	TB	GM	TB	GM	TB	GM
Pseudoword segmentation	**.21***	**.41****	.33**	.65**	.32**	.27*	.39**	.32*
Vowel replacement	.26*	.36**	**.58****	**.72****	.26*	.22	.41**	.25
Word remainder determination	.32**	.45**	.50**	.59**	**.51****	**.33***	.46**	.39**
Sound categorization	.29*	.18	.30**	.28*	.38**	.26*	**.37****	**.34***

Note. Values presented in bold on the diagonal are 6-month stability estimates. Using a Bonferroni approach to control for Type 1 errors across the 8 correlations resulted in a p-value cutoff of .003 to indicate a significant correlation. Coefficients that significantly differed between the German monolingual and Turkish bilingual groups have been enclosed in a box.
*$p < .05$. **$p < .003$ (Bonferroni cutoff).

More importantly, these analyses were conducted to indicate differences between the two groups with regard to scale stability and intercorrelations over time. Using Fisher's z-score transformations of Pearson's r, t-tests were calculated to test for significant differences between longitudinal inter-scale correlations in the two groups. Only one of the 16 pairs of correlations showed a significant difference between groups. Among the monolingual participants, the pseudoword segmentation scale at T1 correlated significantly more strongly with the vowel replacement scale at T2 ($r = .65$) than it did for the bilingual participants ($r = .33$), $t\,(169) = 2.71$, $p = .01$ (two sided). Because the 15 other correlations did not significantly differ between the groups, and because the pseudoword segmentation/vowel replacement correlation discrepancy described above involved highly significant correlations for both groups, it was concluded that the four phonological awareness scales related similarly to one another at both points in time as well as longitudinally in both groups.

In the next step of the investigation of the phonological awareness measures, the validity of using an aggregate scale was examined by way of a traditional factor analyses. Because the items were dichotomous, individual factor analyses for each scale were not possible. Instead, the four scales were analyzed to determine the extent to which they measured a single underlying construct of phonological awareness. These analyses were carried out separately for each group at each time of measurement.

Table 14

Aggregate Phonological Awareness Scale Factor Analyses for each Group at T1 (Middle of 2nd Grade) and T2 (End of 2nd Grade)

	Factor loadings T1		Factor loadings T2	
Scale	Turkish bilingual	German monolingual	Turkish bilingual	German monolingual
Pseudoword segmentation	.62	.74	.47	.71
Vowel replacement	.66	.75	.64	.71
Word remainder determination	.73	.74	.55	.48
Sound categorization	.56	.62	.67	.41
Item variance accounted for	43.6%	50.9%	34.5%	35.0%

The dimensionality of the four subscales from the phonological awareness tests was analyzed using a maximum likelihood factor analysis (see Table 14). Two criteria were used to determine the number of factors to rotate: an a priori hypothesis that four subscales made up a unidimensional instrument and a scree test. For both groups at both points in time, the scree plot clearly indicated that the phonological awareness scales measured a unidimensional construct. As expected from the higher inter-scale correlations at T1 compared to T2 above, the phonological awareness factor at T1 accounted for somewhat more of the item variance at T1 (43.6% for the bilingual group, 50.9% for the monolingual group) than at T2 (34.5% for the bilingual group, 35.0% for the monolingual group). The most important implication, however, is that the factors account for a similar amount of item variance for both groups at each time of measurement. It will therefore be assumed that the factor analytic properties of the aggregate phonological scale, when all scales are taken together, are essentially the same for both the Turkish bilingual and the German monolingual children and viable for use in this study.

Since the verbal pseudoword memory scan was comprised of broader response scales (0 - 3 possible points per item), it was possible to conduct an item-based maximum likelihood factor analysis to explore the dimensionality of the 12 items from the short-term memory measure. Again, the analyses were conducted for both groups separately to control for construct discrepancies between the two groups on the memory span scale. The same two criteria as described above were used to determine the number of factors to rotate (an a priori hypothesis regarding unidimensionality and a scree test). This time, the scree plot indicated that the original hypothesis of unidimensionality was incorrect. Consequently, two factors were rotated using a Varimax rotation procedure. The rotated solution, as shown in Table 15, yielded two interpretable factors in each group, memory for shorter verbal stimuli and memory for more complex verbal stimuli.

Although the same two factors emerged for both groups, some minor differences in factor loadings were detected. For example, the first two-syllable item loaded on both factors for the German monolingual group, but only on the shorter stimuli factor for the Turkish bilingual group. Similarly, the third two-syllable item loaded onto both factors for the bilingual group, but only on the shorter stimuli factor for the monolingual group. The most notable difference between the two groups was the somewhat disparate factor loading of the first three-syllable item, which loaded onto the complex stimuli factor for the bilingual group, but onto the simple stimuli factor for the monolingual group. However, since the factor loadings onto the simple stimuli factor were almost identical for the two groups (monolingual, .36 and bilingual, .35) the loading discrepancy was not considered dramatic. Furthermore, the two factors accounted for an almost identical proportion of the item variance individually and cumulatively (41.8 % for the German monolingual group and 42.0% for the Turkish bilingual group). In essence, the pseudoword memory span measure was determined to have a similar enough factor analytic structure to compare performance of the two groups in further analyses.

Table 15

Verbal Pseudoword Memory Span Factor Analyses for each Group at T1 (Middle of 2nd Grade)

	Factor loadings			
Item	Turkish bilingual		German monolingual	
	Shorter stimuli	Complex stimuli	Shorter stimuli	Complex stimuli
Single syllable item 1	**.51**	.08	**.39**	.15
Single syllable item 2	**.24**	.09	**.42**	.17
Single syllable item 3	**.56**	.16	**.48**	.02
Single syllable item 4	**.76**	.04	**.53**	.14
Two-syllable item 1	**.51**	.23	.46	.41
Two-syllable item 2	**.47**	.23	**.74**	.21
Two-syllable item 3	.40	.44	**.67**	.25
Two-syllable item 4	**.56**	.32	**.73**	.23
Three-syllable item 1	.35	**.50**	**.36**	.19
Three-syllable item 2	.26	**.86**	.14	**.56**
Three-syllable item 3	.18	**.81**	.15	**.99**
Three-syllable item 4	.08	**.74**	.30	**.67**
Item variance accounted for	20.0%	22.0%	23.7%	18.1%

German language scales

The selected measure for assessing expressive vocabulary skills and German language abilities was shortened and utilized for the first time as a research instrument in Germany with Turkish-German bilingual children. It was therefore deemed necessary to thoroughly evaluate the internal psychometric properties of the three bilingual verbal abilities scales (picture vocabulary, synonyms, and antonyms) within this new context. The German listening comprehension measure is also evaluated in this section. Because the Turkish scales are not used as central measures in the analyses to follow, they are not scrutinized here to the extent that the German scales are. With a similar process and sequence as taken with the phonological measures above, the intercorrelations of scales at both times of measurement are considered, followed by a factor analysis, and then a few analyses aimed at determining the scales' external validity.

The first set of analyses calculated concurrent correlations for each group at each time of measurement with Pearson product-moment correlation coefficients to investigate the extent to which the scales correlate with each other and the possibility of disparate correlations among the monolingual and the bilingual groups. Table 16 shows the inter-scale correlations at T1 below the dashed lines (the diagonal) and the T2 correlations above the diagonal for the German monolingual group. All correlations were high and significant.

Table 16

Concurrent Correlations German Vocabulary Subscales at T1 and T2 for the German monolingual (GM) Group: Correlations at T1 are shown below the diagonal; correlations at T2 are shown above the diagonal.

	Picture vocabulary	Synonyms	Antonyms (T2 only)
Picture vocabulary	--	.44**	.58**
Synonyms	.52**	--	.54**

Note. Using a Bonferroni approach to control for Type 1 errors across the 6 correlations resulted in a *p*-value cutoff of .008 to indicate a significant correlation.
**$p < .008$ (Bonferroni cutoff).

An essentially identical pattern emerged in the analysis of concurrent correlations for the Turkish bilingual group shown in Table 17. Correlations between scales at each time of measurement ranged from $r = .35$ to $r = .48$, which was slightly lower than those for the monolingual group were. Nonetheless, there were no meaningful differences between the two groups in the levels of inter-scale correlation at either T1 or T2.

Table 17

Concurrent Correlations German Vocabulary Subscales at T1 and T2 for the Turkish Bilingual (TB) Group: Correlations at T1 are shown below the diagonal; correlations at T2 are shown above the diagonal.

	Picture vocabulary	Synonyms	Antonyms (T2 only)
Picture vocabulary	--	.35**	.48**
Synonyms	.42**	--	.46**

Note. Using a Bonferroni approach to control for Type 1 errors across the 6 correlations resulted in a *p*-value cutoff of .008 to indicate a significant correlation.
$^{**}p < .008$ (Bonferroni cutoff).

Table 18 shows the correlations among the three German vocabulary subscales over the two points of measurement. For the picture vocabulary and synonym scales, which were administered in both mid and late second grade, six-month stability estimates are also available. All longitudinal correlation coefficients were significant at a $p < .008$ level, thus indicating a high level of inter-correlation among the scales even with a half-year period between measurement points and demonstrating high stability over time.

Table 18

Longitudinal Correlations among German Vocabulary Subscales at T1 and T2 for the German Monolingual (GM) Group

	Subscales at T2		
Subscales at T1	Picture vocabulary	Synonyms	Antonyms (T2 only)
Picture vocabulary	**.73****	.50**	.50**
Synonyms	.31*	**.61****	.39**

Note. Values shown in bold on the diagonal are 6 month stability estimates. Using a Bonferroni approach to control for Type 1 errors across the 6 correlations resulted in a *p*-value cutoff of .008 to indicate a significant correlation
$^{*}p < .05$, $^{**}p < .008$ (Bonferroni cutoff).

This was also the case for the longitudinal correlations in the Turkish bilingual group (Table 19). Again, six-month stability was high and all scales correlated with each other significantly from the middle of second grade to the end of second grade. Although there may be a tendency for the longitudinal relationship between the synonym and picture vocabulary scales to be stronger amongst children in the German monolingual group, the longitudinal inter-scale correlation coefficients of the bilingual and monolingual children showed no statistically significant differences. Since this was the case both for the concurrent correlations and the longitudinal scale correlations, it was assumed that for both groups, the three German expressive vocabulary scales functioned and related to one another in sufficiently similar ways.

Table 19

Longitudinal Correlations among German Vocabulary Subscales at T1 and T2 for the Turkish Bilingual (TB) Group

	Subscales at T2		
Subscales at T1	Picture vocabulary	Synonyms	Antonyms (T2 only)
Picture vocabulary	**.70****	.28**	.46**
Synonyms	.40**	**.61****	.36**

Note. Values shown in bold on the diagonal are 6 month stability estimates. Using a Bonferroni approach to control for Type 1 errors across the 6 correlations resulted in a *p*-value cutoff of .008 to indicate a significant correlation
**$p < .008$ (Bonferroni cutoff).

Only at the end of second grade were all three measures administered, thus precluding a factor analysis for the two scales administered in the middle of second grade due to the insufficient number of items. Therefore, a maximum likelihood factor analysis, shown in Table 20, was conducted to analyze the dimensionality of the three vocabulary subtests at T2 only. Again, as was the procedure for the phonological scales, both the a priori hypothesis of unidimensionality and a scree test were used to determine the number of factors. As expected, the scree plot indicated that there was indeed one primary factor being measured. Because the inter-scale correlations above were generally lower in the Turkish bilingual group, it was not surprising that the amount of item variance accounted for was slightly lower for the Turkish bilingual group (44.3%) compared to the German monolingual group (53.0%). Overall, the factor loadings in the two groups did not differ by more than .08, therefore demonstrating strong similarities in scale construction. The aggregate vocabulary scale, which will be used for the principal analyses in this paper, can therefore be regarded as equal in factor analytic properties and scale structure for both groups, and a fair measure for comparison of the two groups.

Table 20

Aggregate German Vocabulary Scale Factor Analyses for each Group at T2 (End of 2nd Grade)

	Factor loadings T2	
Scale	Turkish bilingual	German monolingual
Picture vocabulary	.61	.69
Synonyms	.58	.65
Antonyms	.79	.84
Explained item variance	44.3%	53.0%

Because both the item coding and the length of the Knuspel Listening Comprehension scale (Knuspels Leseverständnis; Marx, 1998) were modified for the current study, a maximum likelihood factor analysis was conducted to verify the measurement of a singular listening comprehension factor with the 11 items. Unfortunately, the authors of the Kunspel reading test did not provide factor analytic data in the test handbook with which the scale validity for this sample could be compared. However, since the authors did not describe the Knuspel Listening Comprehension items as measuring multiple facets of listening comprehension, an a priori hypothesis of unidimensionality was selected. Although five factors emerged with eigenvalues over 1.0 and the scree plot showed no precipitous descent in slope, the first factor appeared relatively stronger than the others did.

As shown in Table 21, the results are dissatisfactory and non-indicative of a unidimensional measure. Not only did the items correlate with the factor erratically and inconsistently within the groups, but they also loaded disparately between the groups. Furthermore, the combination of items explained only a small proportion of the item variance: 13.3% for the Turkish bilingual group and 10.8% for the German monolingual group.

It is not clear why a well-standardized instrument functioned so poorly in this context. There are several possible explanations, though. First, it is possible that the modification of the instrument with the collapsed coding of instruction comprehension and question comprehension points for each item (as originally intended by the authors) confounded the factors (see the measure description in Section 3.3.3.). Second, this instrument may not be suited for use as a research instrument, but instead, more appropriate as a purely diagnostic measure for teachers in classroom settings. Third, since the factor analytic properties of the instrument were not found in the author's description of the instrument, it could be inferred that the authors neither developed the scale factor analytically nor tested it with a factor analysis for the standardization sample. In other words, the items of the Knuspel listening comprehension scale may not be measuring the child's underlying ability for listening comprehension at all. Therefore, this instrument is used only sparingly and for explorative purposes, not for principal analyses in this report.

Table 21

German Listening Comprehension Factor Analyses for each Group at T2 (End of 2nd Grade)

	Factor loadings	
	Turkish bilingual	German monolingual
Item a	.52	.34
Item 1	.27	.37
Item 2	.43	.45
Item 3	.22	.08
Item 4	.16	.18
Item 5	.12	.10
Item 7	.66	.16
Item 8	.41	.13
Item 9	.41	.41
Item 11	.29	.60
Item 12	-.02	.36
Explained item variance	13.3%	10.8%

Correlations with external measures and background variables

This section aims to examine the convergent validity of the utilized measures and to investigate the general correlations between all principal scales. In addition, this section explores significant differences in correlative patterns. First, comparisons of the individual German language scales and the teachers' assessments of German proficiency provide indications of the external validity of the various scales. Second, correlations between the German language scales and the personal background information provided by the participants allow for both the exploration of the relationship between out-of-school experiences and language skills as well as external scale validity. Finally, an overall principal measures correlation table shows the scale intercorrelations for the Turkish bilingual and German monolingual groups separately. These analyses should contribute toward a more multifaceted comprehension of the measures utilized, an awareness of potential influences on those measures, and an indication of any possible discrepancies between the two groups.

Table 22

Correlations between Teacher Assessments of German Language Abilities at T1 and German Vocabulary Scales for each Group

	Teacher assessment of German language abilities mid-2nd grade (T1)	
	TB	GM
Aggregate vocabulary T1	.57**	.40**
Picture vocabulary	.49**	.36**
Synonyms	.51**	.34*
Aggregate vocabulary T2	.47**	.45**
Picture vocabulary	.42**	.33*
Synonyms	.24*	.34*
Antonyms	.42**	.46**
German listening comprehension T2	.28*	.35**

Note. TB = Turkish bilingual, GM = German monolingual. Using a Bonferroni approach to control for Type 1 errors across the 8 correlations resulted in a *p*-value cutoff of .006 to indicate a significant correlation.
**p* < .05, ***p* < .006 (Bonferroni cutoff).

Table 22 provides an indication of the convergent validity of the German language measures by comparing the vocabulary scales from the middle (T1) and end (T2) of second grade and the listening comprehension measure from the end of the second grade with the teachers' reported perception of their individual students' German language abilities in the middle of second grade (T1). In general, the teachers' ratings correlated strongly with the aggregate vocabulary scales at both T1 and T2 for the bilingual group ($r = .57$ and $r = .47$, respectively) as well as for the monolingual group ($r = .40$ and $r = .45$, respectively). Correlations with the individual scales were slightly lower, but not statistically. Scores on the listening comprehension measure were also moderately correlated with the teacher assessments (TB: $r = .28$, GM: $r = .35$) albeit more weakly than the better substantiated measures of German vocabulary (at T1, TB: $r = .57$, GM: $r = .40$; at T2, TB: $r = .47$, GM: $r = .45$). Overall, these analyses indicate an equally high convergent validity for the German vocabulary scales for both the Turkish bilingual and German monolingual groups.

As discussed in detail in Section 1.2.4., the effects of home environment, childcare experiences, and class composition are thought to have an important influence on language development in children, especially those with minority language backgrounds. To investigate the plausibility of such hypotheses in this sample, and to look for possible differences between the bilingual and monolingual children, several variables from the home environment (number of siblings, number of older siblings, parent reading behavior) and several institutional variables (preschool attendance, after-school daycare) were examined in relation to German vocabulary performance.

Table 23 presents the results for the German monolingual group. Among the monolingual participants, the background variables were unrelated to the German language measures. The only exception to this was the positive correlation between after-school daycare and German listening comprehension.

Table 23

Correlations between Background Information and German Language Scales for the German Monolingual (GM) Group

	Home environment			Educational environment	
	Number of siblings	Number of older siblings	Parent reading aloud behavior	Preschool attendance	After-school daycare attendance
German vocabulary T1	-.08	-.05	.19	.17	.03
German vocabulary T2	-.14	-.09	-.14	-.08	.21
German listening comprehension	-.20	-.21	.07	-.08	.29*

Note. Using a Bonferroni approach to control for Type 1 errors across the 18 correlations resulted in a *p*-value cutoff of .003 to indicate a significant correlation. No correlations in this analysis reached that level of significance.
*$p < .05$.

A similar analysis of background information and language skills was conducted for the bilingual group, with two small changes. First, the Turkish vocabulary scales were included in the analysis and second, the amount of Turkish spoken in the home was included as one of the home variables. The correlations between the German language measures and home variables displayed in Table 24 were all negligible, ranging from -.19 to .04. The only home environment variable that implied a minimally significant association was the amount of Turkish spoken in home scale in relation to the Turkish vocabulary scale at T2 ($r = .20$).

The institutional educational variables however, produced some significant correlation coefficients. A moderately significant correlation emerged between German language skills and the attendance at an after-school daycare institution: after-school daycare attendance was positively related with German vocabulary at the end of second grade (T2; $r = .23$). It should be pointed out, however, that did not hold true at T1 and should therefore be interpreted with caution. Conversely, attending an after-school daycare program appeared to be significantly negatively related to Turkish vocabulary ($r = -.29$) at a $p < .01$ level.

Table 24

Correlations between Background Information and German Language Scales for the Turkish Bilingual (TB) Group

	Home environment				Educational environment	
	Number of siblings	Number of older siblings	Parent reading aloud behavior	Amount of Turkish language in home	Preschool attendance	After-school daycare attendance
German vocabulary T1	.03	.00	.04	-.19	.08	-.04
German vocabulary T2	-.06	-.12	-.11	-.09	.17	.23*
German listening comprehension	-.08	-.05	-.13	-.10	.23*	.01
Turkish vocabulary T1	-.02	.03	.11	-.06	-.13	-.29*
Turkish vocabulary T2	-.05	-.10	-.07	.20*	-.01	-.12

Note. Using a Bonferroni approach to control for Type 1 errors across the 30 correlations resulted in a *p*-value cutoff of .001 to indicate a significant correlation. No correlations in this analysis reached that level of significance.
$^*p < .05$.

Generally, it should be kept in mind that the available information regarding the home environment variables were based on a self-report instrument administered to young children. It may therefore be lacking in validity. As for the institutional variables, it seems that attending after-school daycare might be advantageous for the language skills of children in both groups. Still, no significant differences were detected between the two groups in the relationships between their German language skills and external home or institutional variables. This can be taken as another positive indication that the expressive vocabulary instruments are measuring similar underlying constructs in both groups with equal validity.

The following analyses provide a comprehensive summary of all essential scale intercorrelations at all times of measurement for the German monolingual (Table 25) and the Turkish bilingual (Table 26) groups. These tables provide the best initial overview of how the verbal and reading abilities fundamental to this investigation correlated with each other. The left side displays the intercorrelations between the verbal measures. Correlations on the right side represent the correlation coefficients for the verbal and written measures. Listening comprehension was included as an exploratory variable, but will not be discussed at length here.

For the monolingual group, the verbal measures (excluding listening comprehension) significantly correlated with each other at a $p < .05$ level. The six-month stabilities of the aggregate phonological awareness scales and the German vocabulary scales were particularly high ($r = .73$ and $r = .74$, respectively). On the right hand side of the table, we see that verbal memory is not significantly correlated with cognitive abilities[1], but more strongly related to the measures of reading with correlation coefficients ranging from $r = .20$ to $r = .33$. The strongest correlations with the written measures were found for the phonological awareness scales at both T1 and T2. Phonological awareness in second grade was significantly related to the non-verbal measure of cognitive ability in first grade (T-1) with an $r = .43$ at T1 but decreasing to $r = .32$ at T2. The phonological awareness scales correlated with the reading measures with coefficients ranging from .31 to .71 at both times of measurement and a mean correlation with reading measures at T1 of .48 and .56 at T2. Compared to the bilingual group, the vocabulary measures were less strongly correlated with the reading measures for the German monolingual group at both T1 and T2 (range: $r = .20$ - .48). Phonological awareness and vocabulary correlated similarly with cognitive abilities with correlations around .40.

In the Turkish bilingual group, the verbal measures were all strongly intercorrelated. The six-month stability of the aggregate phonological and vocabulary measures were also moderately high (.68 and .66, respectively). Similar to the German monolingual group, verbal short-term memory correlated insignificantly with cognitive abilities and moderately with the reading measures (range: .20 - .31). Also congruent to the monolingual group, the reading measures produced the highest correlations with the phonological awareness measures at both times of measurement. The average correlation between phonological awareness at T1 and the reading measures over time was $r = .52$ and $r = .43$ at T2. Again, as in the monolingual group, the

[1] The use of a non-verbal cognitive abilities test for the linguistically diverse sample at hand is a likely reason for the surprisingly low correlation between verbal memory and cognitive abilities in both groups.

correlations between the vocabulary scales and the reading measures were somewhat lower, ranging between .20 and .48.

With a small number of exceptions, there were few significant differences in these intercorrelations between the two groups. Significant differences were tested with *t*-tests after applying Fisher's *r* to *z* transformation. First, verbal memory correlated significantly higher with German vocabulary at T2 for the German monolingual participants ($r = .38$) than for the Turkish bilingual group ($r = .09$), $t\,(169) = 1.94$, $p < .05$. The second notable difference was found in the correlations between reading comprehension at T3 and phonological awareness at T2. This correlation was much stronger for the monolingual group ($r = .71$) than for the bilingual group ($r = .38$), $t\,(144) = 3.05$, $p < .01$. In general, the phonological awareness measure at T2 tended toward weaker correlations with the written measures for the bilingual group. Finally, there was indication of a significant difference in the extent to which listening comprehension and verbal memory were related, $t\,(169) = 1.97$, $p < .05$. Listening comprehension correlated more closely to the verbal measures within the bilingual group than within the monolingual group.

These correlations, although in essence not substantially different between the two groups, do show some variance in the interplay between verbal abilities and reading competencies. Using a series of multivariate analyses, the possibility of differing mechanisms leading to early reading for German monolingual and Turkish bilingual children will be explored in detail below. Due to its ambiguous factor structure, the listening comprehension measure will be used sparingly. However, the remaining analyses above substantiated the construct validity of the verbal memory, phonological awareness, and expressive vocabulary measures in both subscale and aggregate form at all times of measurement. These measures will therefore be used as the principal variables for addressing the research questions at hand.

Table 25

Complete Intercorrelations for all Central Measures at each Time of Measurement for the German Monolingual (GM) Group

	Verbal measures						Written measures						
	Verbal memory	Phonological awareness		German vocabulary		Listening comprehension	Cognitive abilities	Word decoding				Reading comprehension	
	(T1)	T1	T2	T1	T2	(T2)	(T -1)	T0	T1	T2	T3	T2	T3
Verbal short-term memory (T1)	--						.14	.29*	.33**	.29*	.25	.20	.31*
Phonological awareness													
T1	.38**	--					.43*	.56**	.47**	.49**	.32*	.47**	.57**
T2	.39**	.73**	--				.32*	.59**	.53**	.60**	.40**	.56**	**.71****
German vocabulary													
T1	.31*	.42**	.43**	--			.41*	.39**	.31**	.24*	.20	.26*	.41**
T2	**.38****	.38**	.54**	.74**	--		.42*	.46**	.34**	.30*	.24	.35*	.48**
Listening comprehension (T2)	**.04**	.25*	.28*	.15	.28**	--	.18	.18	.11	.17	-.01	.19	.13

Note. Coefficients that significantly differed between the German monolingual and Turkish bilingual groups have been marked bold.
$^{*}p < .05$. $^{**}p < .01$.

Table 26

Complete Correlation Coefficients for all Central Measures at each Time of Measurement for the Turkish Bilingual (TB) Group

	Verbal measures						Written measures						
	Verbal memory	Phonological awareness		German vocabulary		Listening comprehension	Cognitive abilities	Word decoding				Reading comprehension	
	(T1)	T1	T2	T1	T2	(T2)	(T -1)	T0	T1	T2	T3	T2	T3
Verbal short-term memory (T1)	--						.20	.24*	.31**	.23*	.32**	.20	.23*
Phonological awareness													
T1	.54**	--					.33**	.53**	.56**	.51*	.46**	.49**	.55**
T2	.32**	.68**	--				.30**	.44**	.49**	.50*	.36**	.39**	**.38****
German vocabulary													
T1	.15	.45**	.42**	--			.28**	.18	.34**	.35**	.32**	.47**	.36**
T2	**.09**	.36**	.46**	.66**	--		.23*	.22*	.32**	.32**	.29**	.38**	.34**
Listening comprehension (T2)	**.34****	.39**	.32**	.39**	.41**	--	.22*	.19	.25*	.25*	.27*	.32**	.35**

Note. Coefficients that significantly differed between the German monolingual and Turkish bilingual groups have been marked bold.
*p < .05. **p < .01.

4.1.2. Participant language and learning patterns

This section describes the language and literacy behaviors of the participants in order to present a multi-faceted picture of the children's linguistic experiences beyond the performance measures administered within the study. The analyses in this section allow for the comparison of the two groups at the onset of formal education on language and learning assessments. Information regarding the participants' observed language and learning behaviors in the classroom was gleaned from teacher assessments in early Grade 1 (T-1) and mid-Grade 2 (T1; see Section 2.2. for an overview of the times of measurement). The German monolingual and Turkish bilingual groups were compared with regard to their language and behavior in school to investigate possible differences between the groups.

Furthermore, the correlational analyses of both Turkish and German language abilities provide important insights into the interplay between the bilingual children's two languages. Although often neglected in studies of this nature, it is essential to have as much information as possible regarding the linguistic profiles of the bilingual sample with regard to their dual language development. Special attention is paid to the language development of the bilingual participants through the second grade (T1 to T2).

Teacher assessments

Raw scores and significance tests for each of the three teacher assessment scales can be found in Table 27. At both the beginning of the first grade and in the middle of second grade, teachers estimated the German language abilities of the bilingual group to be significantly poorer than that of the monolingual group, $F(1, 156) = 42.99, p < .01$, and $F(1, 152) = 22.78, p < .01$ respectively. However, the teachers perceived a higher level of learning readiness by the bilingual group by the middle of second grade, $F(1, 152) = 4.73, p = .03$. Teachers reported no differences between the two groups' abilities to concentrate in class. In sum, the teachers noted significant differences in language abilities among participants in the two groups and a certain tendency for children in the bilingual group to show higher levels of learning readiness (in mid-second grade only), but no differences in the participants' ability to focus their attention in class.

Table 27

Teacher Assessments of Language and Learning abilities at T-1 and T1 (1 = Very Poor, 4 = Excellent)

		GM	TB	*F*	*p*
German language abilities: Early Grade 1					
	M	3.6	2.9	42.99	.00
	SD	0.53	0.71		
German language abilities: Mid-Grade 2					
	M	2.44	2.09	22.78	.00
	SD	0.43	0.46		
Readiness to learn: Early Grade 1					
	M	2.88	2.97	1.28	.26
	SD	0.55	0.51		
Readiness to learn: Mid-Grade 2					
	M	2.80	3.00	4.73	.03
	SD	0.59	0.55		
Concentration abilities: Early Grade 1					
	M	2.69	2.73	0.15	.69
	SD	0.66	0.57		
Concentration abilities: Mid-Grade 2					
	M	2.65	2.70	0.22	.64
	SD	0.75	0.54		

Note. GM = German Monolingual, TB = Turkish Bilingual.

Bilingual Turkish and German vocabulary abilities

The following three analyses aim to illustrate the Turkish and German verbal abilities of participants in the bilingual group. Table 28 shows the relative dominance of German over Turkish with regard to expressive vocabulary skills and development[1]. Since comparing translated scales in two languages is a questionable process due to the inevitable uncertainty of exact parallel psychometric properties and cultural constructs (e.g., see Geisinger, 1994) only raw scores are compared here. By examining the raw scores of correctly identified words, it is apparent that the bilingual group identified substantially more words in German than in Turkish for almost every subtest at every time of measurement. For example, by the end of second grade, the bilingual children verbally identified almost twice as many picture stimuli in German (*M* = 11.89) than they were able to identify in Turkish (*M* = 5.93). This tendency is also clear in the

[2] Raw scores are based on total items administered for each scale without the removal of any items for the purpose of estimating score development over time.

aggregate vocabulary scales at T2 on which the bilingual children identified 18.94 words on average in Turkish but 26.09 words on average in German.

Table 28

A Comparison of Turkish and German Vocabulary Skills and Development among Participants in the Turkish Bilingual Group from the Middle of 2nd Grade (T1) to the End of 2nd Grade (T2)

	T1		T2		Development in Turkish from T1 to T2		Development in German from T1 to T2	
	Turkish	German	Turkish	German	*t*	*p*	*t*	*p*
Picture vocabulary								
M	6.64	10.59	5.93	11.89	-2.26	.03	3.70	< .01
SD	3.07	4.66	2.64	4.55				
Synonyms								
M	5.46	3.59	5.60	4.27	0.71	.48	3.00	< .01
SD	1.64	2.22	1.98	2.59				
Antonyms[a]								
M	--	--	7.41	9.91	--	--	--	--
SD			2.89	2.48				
Aggregate vocabulary score[a]								
M	11.94	14.42	18.94	26.09	--	--	--	--
SD	4.48	6.16	6.70	7.74				

[a] The antonym scale was administered only at T2; the inclusion of the antonym scale into the aggregate score at T2 makes it impossible to measure the growth in either scale.

In general, the discrepancy between Turkish and German vocabularies of the bilingual participants seems to grow over time. While the difference between the two aggregate scales at T1 was only 2.48, there was a discrepancy of 7.48 words a half-year later. The only scale on which the bilingual participants' German vocabularies did not appear stronger than their Turkish vocabularies was on the synonym scale. In fact, at both times of measurement the participants identified more synonyms in Turkish. But again, this discrepancy shifted over time with the bilingual children's German vocabulary skills increasing significantly between T1 and T2 (for picture vocabulary: $t(100) = 3.70, p < .01$ and for synonyms: $t(100) = 3.00, p < .01$). Conversely, their ability to identify Turkish words seemed to decrease between T1 and T2 for the picture vocabulary task: $t(100) = -2.26, p = .03$, and stagnate for synonym identification: $t(100) = 0.71, p = .48$. Overall, it appears that after one to two years of instruction in German within the public education system, the Turkish bilingual children in this sample are not dominant in Turkish and that their German skills seem to be gaining dominance over their Turkish skills with regard to academic vocabulary.

Although not an explicit research question in this study, this data set allows for the exploration of the relationship between German and Turkish vocabulary skills among bilingual Turkish-German speakers. The next set of analyses explores the question of language interdependence (e.g., Cummins, 1979; see Section 1.2.2. for a description) among bilingual children in order to obtain a clearer picture of the bilingual participant's linguistic abilities. Furthermore, this affords the rare opportunity to address the commonly asked question as to what extent the L1 and L2s of minority language children are related (e.g., Verhoeven, 1996). Pearson product-moment correlation coefficients were computed among the Turkish and German expressive vocabulary scales as well as the cognitive abilities measure for participants in the bilingual group. Table 29 shows that several correlations were found of small to moderate sizes; 9 of the 22 correlations were found to be significant at a .05 level. There appears to be a pattern of significant correlations on the diagonal of Table 29: The corresponding scales in Turkish and German indicated the highest correlations. The largest scale correlations were found between synonym identification in both languages at both T1 and T2 ($r = .33$ and $r = .31$, respectively). Cognitive abilities measured in the middle of first grade correlated similarly with vocabulary in both languages, with moderate correlations at T1 for picture vocabulary identification but no meaningful correlations between cognitive abilities at T-1 and expressive vocabulary by the end of second grade.[2]

With the exception of tasks involving synonym identification, there were few strong correlations between vocabulary skills in the bilingual children's first and second languages. Non-verbal cognitive skills seem to play a similarly minor role in the vocabulary abilities in both languages as well. The implications of these findings for the interdependence hypothesis will be addressed in the Discussion section of this paper.

[2] Calculating the partial correlations between Turkish and German vocabulary abilities by controlling for cognitive abilities produced essentially identical results as the bivariate correlations including cognitive abilities as shown in Table 29. No additional significant correlations emerged as a result of controlling for cognitive abilities.

Table 29

Correlations between Turkish and German Versions of the Vocabulary Measures for the Bilingual Group with Consideration for Cognitive Abilities

	Cognitive abilities T -1	German picture vocabulary T1	German synonyms T1	German picture vocabulary T2	German synonyms T2	German antonyms T2
Cognitive abilities T -1	--	.30*	.15	.17	.19	.20
Turkish picture vocabulary T1	.29*	.26*	.16	.02	.04	.08
Turkish synonyms T1	.23*	.19	.33**	.01	.29*	.01
Turkish picture vocabulary T2	.11	-.03	.05	-.05	.05	.13
Turkish synonyms T2	-.03	.03	.11	.00	.31**	.25*
Turkish antonyms T2	.07	-.08	.05	-.10	.14	.30*

Note. Using a Bonferroni approach to control for Type 1 errors across the 22 correlations resulted in a p-value cutoff of .002 to indicate a significant correlation.
$*p < .05$, $**p < .002$ (Bonferroni cutoff).

As a continuation of the interdependence analyses, Table 30 presents information regarding the extent to which Turkish and German expressive vocabulary skills are associated with performance on measures of cognitive abilities, phonological awareness, decoding, and reading comprehension for children in the bilingual group. Turkish expressive vocabulary measured in the middle of second grade seems to correlate moderately with non-verbal cognitive abilities measured in the middle of first grade ($r = .28$), but no other scales correlated with Turkish vocabulary skills at either time of measurement. In contrast, the bilingual children's German skills correlated moderately to highly with the base reading and reading measures at both times of measurement with correlation coefficients ranging from .29 to .47. The measure of cognitive abilities at T-1 correlated with Turkish vocabulary only at T1, but with German vocabulary at both T1 and T2. The scale with which German vocabulary seemed to be most closely associated was phonological awareness with an average correlation coefficient of .42 over all points of measurement. In general, it can be assumed that German academic vocabulary skills are more strongly associated with phonological awareness, German word decoding, and German reading than are Turkish academic vocabulary skills.

Table 30

Differential Correlations of Turkish and German Versions of the Vocabulary Measures with all (German Language) Reading Tasks for the Bilingual Group

	Cognitive abilities T -1	PA T1	PA T2	Decode T2	Decode T3	Reading T2	Reading T3
Turkish vocabulary aggregate T1	.28*	.18	.14	.04	.17	.07	.10
Turkish vocabulary aggregate T2	.10	.11	-.02	.02	.20	-.03	-.16
German vocabulary aggregate T1	.28*	.45**	.42**	.35**	.47**	.36**	.36**
German vocabulary aggregate T2	.23*	.46**	.36**	.32**	.29*	.38**	.34**

Note. PA = Phonological awareness, Decode = German word decoding, Reading = German reading comprehension. Using a Bonferroni approach to control for Type 1 errors across the 22 correlations resulted in a *p*-value cutoff of .002 to indicate a significant correlation.
*p < .05. ** p < .002 (Bonferroni cutoff).

4.2. Mean differences between bilingual and monolingual readers

This first set of hypotheses proposes a series of mean differences between the Turkish bilingual and German monolingual participants. The hypotheses that Turkish bilingual students would perform better on measures of phonological awareness and verbal memory than their German monolingual peers (H1a), that the bilingual group would perform significantly more poorly on measures of German vocabulary (H1b), and that the groups will not differ with regard to their abilities to decode words (H1c) are tested in this section. One-way multivariate analyses of variance with covariates (MANCOVAs) were calculated at each time of measurement for H1a and H1b to evaluate whether the population means on sets of dependent variables (here, the instrument subtests) vary across the groups. The MANCOVA procedure was deemed appropriate for measuring group differences because it tests the hypothesis that the population means for multiple dependent variables are the same for all levels of a factor (across the two groups), while taking covariates into account. The MANCOVA not only evaluates the equality among groups on a series of dependent variables, but also tests the equality among the groups on linear combinations of the dependent variables (see Green et al., 2000). Since the instrument measuring decoding did not have subtests to be investigated multivariately, a one-way repeated measures analysis of variance was conducted controlling for covariates over the four points of measurement. All three analyses integrated gender and cognitive abilities as covariates.

H1a: Bilingual Turkish children will perform better on most of the measures of phonological awareness and verbal memory than their German counterparts

One-way MANCOVAs for each time of measurement examined whether inequalities existed between Turkish bilingual and German monolingual second graders with regard to their performance on the four phonological awareness subtests and the pseudoword short-term verbal memory test. Results of the analyses for the measures at the first time of measurement (T1: middle second grade) and at the second time of measurement (T2: end of second grade), along with the means and standard deviations for each group are found in Table 31.

Two preliminary procedures were carried out to ensure that integral assumptions of both multivariate analyses of variance and covariance were fulfilled. To ascertain if the population variances and covariances among the dependent variables were the same across both groups (a central assumption underlying the one-way MANOVA), the homogeneity of the variance-covariance matrices was tested with Box's M statistic for the dependent variables at both T1 and T2. The F test was non-significant at both times of measurement, F (15, 57017) = 1.32, p = .18 at T1 and F (10, 66486) = 1.02, p = .43 at T2. The dispersion matrices for both Turkish bilinguals and German monolinguals are therefore relatively homogenous with regard to the subscales of the phonological awareness tests at both times of measurement. The homogeneity-of-slopes assumption, fundamental to analyses of covariance, was tested by evaluating the interaction between the two covariates and group membership in predicting the dependent variables. The preliminary analyses evaluating this assumption indicated that the relationships between the covariates (gender and cognitive abilities) and the dependent variables did not differ significantly as a function of group membership (bilingual or monolingual) at either T1 or T2. Appendix C provides a table (Table C4) showing the Wilks Λ and F values of those tests. The slopes can therefore be assumed to be homogeneous and the analysis of covariance appropriate for the sample and measures.

Table 31

Means and Standard Deviations for the Phonological Awareness Scale at T1 and T2 for the Turkish Bilingual (TB) and German Monolingual (GM) Groups with F and η^2 Values, Controlling for Gender and Cognitive Abilities

	M				
	TB	GM	*df*	*F*	Partial η^2
Time 1					
Pseudoword segmentation	4.60	4.11	1	2.10	.01
SD	2.47	2.18			
Vowel replacement	4.88	4.07	1	2.21	.02
SD	4.12	3.85			
Word remainder determination	5.63	5.28	1	1.93	.01
SD	1.81	2.02			
Sound categorization	4.74	4.96	1	0.70	.01
SD	2.18	2.33			
Verbal memory	13.02	11.54	1	2.47	.02
SD	6.12	6.18			
Time 2					
Pseudoword segmentation	4.44	3.80	1	5.43*	.04
SD	2.00	1.94			
Vowel replacement	6.46	6.13	1	0.21	.00
SD	4.35	4.24			
Word remainder determination	6.29	5.93	1	1.38	.01
SD	1.30	1.51			
Sound categorization	4.42	3.97	1	1.16	.01
SD	1.97	2.29			

*$p < .05$.

Despite the fact that, with the exception of the sound categorization subtest at T1, the Turkish bilingual group performed consistently better on all subtests at both times of measurement, no significant differences were found between the groups on the dependent measures at either T1, Wilks $\Lambda = .97$, F (5, 139) = .79, $p = .56$, $\eta^2 = .03$; or at T2, Wilks $\Lambda = .96$, F (4, 140) = 1.53, $p = .20$, $\eta^2 = .04$. Analyses of covariances (ANCOVA) were conducted on each dependent variable as follow-up tests. The F values and η^2 for the ANCOVAs as well as the means and standard deviations for each scale are illustrated in Table 31. Although the pseudoword segmentation subscale shows a significant difference between the groups at a $p < .05$

level, using the Bonferroni method to avoid Type 1 errors would call for the dependent variables to be tested at a .0125 level. Under this stringent constraint, no significant differences were detected between the two groups at either time of measurement.

Overall, with regard to H1a, the proposition that Turkish bilingual children would perform better on measures of phonological awareness had to be rejected in favor of the null hypothesis. A tendency for the bilingual children to correctly respond to tasks requiring the analysis and modification of phonological stimuli more often than their German monolingual classmates was apparent (the Turkish bilingual group scored higher on every subtest but sound categorization at T1), but this trend could not be confirmed with strict significance tests within multivariate analyses of variance.

H1b: The Turkish bilingual students will score more poorly on measures of German vocabulary than their monolingual German peers

Identical to the procedure to the above for the phonological scales, a MANCOVA explored the hypothesized mean differences in the German vocabulary scales. As in the procedure described for testing H1a, two preliminary analyses were conducted to ensure that the data met the requirements of 1) a homogeneous variance-covariance matrix for both groups and 2) homogeneity of slopes in the interaction between the covariates and group membership in predicting the dependent variables, in this case, the German vocabulary scales. First, the Box Test for homogeneity of dispersion matrices revealed no significant group differences at either time of measurement, T1 F (3, 620854) = 1.79, p = .15, T2 F (10, 66486) = .92, p = .51. The homogeneity-of-slopes assumption was also found to hold true with non-significant results for both covariates and every dependent variable at each point in time, indicating that the relationship between the covariates and the dependent variables was not a function of group membership[3].

With both underlying assumptions met, two MANCOVAs were conducted to determine the effect of group membership on the two German vocabulary variables at T1 and on the four variables at T2. Significant differences were found between the two groups in favor of the German monolingual group both in the middle of second grade, Wilks Λ = .67, F (2, 142) = 34.35, p < .01, and six months later at the end of second grade Wilks, Λ = .71, F (4, 140) = 14.22, p <.01. The multivariate effect sizes were also substantial, η^2 = .33 at T1 and , η^2 = .29 at T2.

[3] Appendix C provides a table (Table C4) with the Wilks Λ and F values for the tests of slope homogeneity.

Table 32

Means and Standard Deviations for the Vocabulary Scale at T1 and T2 for the Turkish Bilingual (TB) and German Monolingual (GM) Groups with F and η^2 Values, Controlling for Gender and Cognitive Abilities

	M				
	TB	GM	*df*	*F*	Partial η^2
Time 1					
Picture vocabulary	7.79	13.27	1	66.57**	.32
SD	5.03	7.79			
Synonyms	3.53	4.40	1	4.87*	.03
SD	2.15	2.41			
Time 2					
Picture vocabulary	8.93	14.32	1	57.91**	.33
SD	4.54	3.84			
Synonyms	4.17	5.19	1	5.45*	.11
SD	2.54	2.89			
Antonyms	6.08	7.47	1	12.35**	.13
SD	2.40	2.61			
Listening comprehension	18.87	20.05	1	4.10*	.03
SD	3.22	3.00			

Note. Using a Bonferroni approach to control for Type 1 errors for the four dependent resulted in a *p*-value cutoff of .002 to indicate a significant correlation.
$^*p < .05$. $^{**}p < .025$ at T1 $^{**}p < .0125$ at T2, (Bonferroni cutoffs).

Table 32 contains the means, standard deviations, *F* values, and partial η^2 estimates for the ANCOVAs that were conducted as follow-up tests to the MANCOVA. Although all *F* values showed significant differences between the two groups at a $p < .05$ level, using a Bonferroni method to avoid Type 1 errors required the alpha cutoff for the analyses at T1 to be .025 for the two dependent variables, thus leaving only the picture vocabulary subtest clearly significant at T1. For the synonym scale, the $F\,(1, 24) = 4.87$ reached an alpha of only $p = .029$, thus just missing the cutoff. At T2, the Bonferroni method set a significance cutoff at .0125 for the four dependent variables, leaving only the picture vocabulary and antonym scales with significant differences between the two groups.

The MANCOVA analysis at both T1 and T2 clearly supported Hypothesis 1b with regard to the German monolingual group's superior performance on a series of German vocabulary and language measures. Although using stringent constraints for significance levels indicated significant differences in only three of the six ANCOVAs, the trend is clear and confirmed by the

MANCOVA. As predicted, second grade German monolingual children performed significantly better on measures of German vocabulary than did their Turkish bilingual counterparts.

H1c: The Turkish-German bilingual group will not differ from the German monolingual sample on measures of word decoding

Since the measure of decoding was a single scale administered at six-month intervals from the first to the third grade, a one-way between-subjects repeated-measure analysis of variance with two covariates was selected to test mean differences in decoding abilities between the Turkish bilingual and German monolingual groups. Gender and cognitive abilities were controlled as covariates in the analysis, which required the data to again be tested for the homogeneity-of-slope assumption. As with the phonological and German verbal abilities data, the preliminary analysis revealed no significant differences in the relationships between the covariates and the decoding measures as a result of group membership[4]. The inclusion of the two covariates was thus considered acceptable in the repeated measure analysis of variance.

Table 33

Means and Standard Deviations for Word Decoding, from 1st (T0) to 3rd Grade (T3)

		M	*SD*
T0	TB	25.50	14.36
	GM	28.32	13.54
T1	TB	46.01	16.79
	GM	47.54	16.26
T2	TB	59.90	16.45
	GM	62.68	17.08
T3	TB	71.74	17.26
	GM	73.07	19.99

Note. GM = German Monolingual, TB = Turkish Bilingual.

The one-way between-subjects repeated-measure ANCOVA was conducted with the within-subject factor as time, the between-subject factor of group (Turkish bilingual, German monolingual), and the dependent variable of decoding scores over four points in time. Table 33 shows the means and standard deviations for the decoding measure. The results of the ANCOVA indicated a significant time effect, Wilks $\Lambda = .72$, $F\ (3, 108) = 13.73$, $p < .01$,

[4] Appendix C provides a table (Table C4) with the Wilks Λ and F values for the tests of slope homogeneity.

multivariate η^2 = .28, but no interactions or meaningful differences between groups at any of the four times of measurement. This analysis will be discussed in greater depth in section 4.4. with regard to the students' development over the four points of measurement. Its purpose here is to provide substantiation for H1c, that the German monolingual and Turkish bilingual children would not differ in their performance on tasks of decoding. The similar scores of the two groups at each point in time clearly support H1c. In sum, no differences in decoding abilities were found between the German monolingual and Turkish bilingual children in the first, second, or third grades.

In sum, the above analyses substantiated two of the three hypotheses regarding mean differences between the German monolingual and Turkish bilingual participants. MANCOVAs and one-way repeated-measure between-subjects ANCOVA showed that, although at first glance the Turkish bilingual children performed better than their monolingual peers on measures of phonological awareness, this was not confirmed with stringent tests of significance. Acceptance of the null hypothesis that the groups are equal in performance on tasks requiring phonological awareness was therefore necessary. It also appeared that there was enough evidence to reject the null hypothesis and accept H1b which predicted poorer performance among the Turkish bilingual participants compared to the German monolingual participants on measures of German vocabulary. Finally, H1c was also confirmed with findings over four times of measurement indicating no statistical or even suggestive evidence of mean differences between the two groups on measures of decoding abilities from the first to the third grade.

4.3. Differential predictors of reading skills

In this section, the predictors hypothesized to be essential in learning to read are tested with the Turkish bilingual and German monolingual samples. The predictor variables examined in this section are primarily limited to the aggregate phonological awareness scales at T1 and T2, the aggregate German vocabulary scales from T1 and T2, cognitive abilities from T-1, decoding at T1 and T2, and verbal short-term memory at T1. In order to disentangle the factors related to reading comprehension, the predictors of word reading were examined before investigating the predictors of text reading. Therefore, although not a primary hypothesis in this thesis, this section examines the regression weights of the predictors of word decoding at T2 and T3. Reading comprehension at T2 and T3 are, however, the principal criterion variables in this section.

Hypothesis 2 stated that the individual predictors for decoding (H2a) and reading comprehension (H2b) would be essentially the same for both groups. In other words, no variables would be meaningful for one group while thoroughly irrelevant for the other. This was tested with two sets of Pearson product-moment correlations, one for each group. The coefficients were then tested for significant differences between the Turkish-German bilingual and German monolingual children.

In contrast to Hypotheses 2b, 2c, and 2d which are all aimed at predicting reading comprehension, Hypothesis 2a dealt with the relationship between the predictor variables and word reading (decoding). The remaining hypotheses predicted that although Turkish-German bilingual and German monolingual readers share the same component processes required for reading comprehension (H2b), phonological awareness would play a weaker role in predicting reading comprehension for the bilingual children (H2c), while vocabulary skills would play a larger role (H2d) for the bilingual children. Those predictions were separately tested for each group with a series of hierarchical multiple regressions, one for each combination of predictor and criterion variable measurement points (T1/T2, T1/T3, T2/T3). As a follow-up to the separate regressions for each group, the specific hypotheses were investigated by way of including an interaction term into the regression equation to test for significant group differences on specific predictors.

Multiple regression analyses are commonly used to test the predictive strength of phonological skills and other measures for estimating reading proficiency. Examining the differential predictive power of each component related to reading can create a clearer picture of which mechanisms lead to successful reading abilities in two linguistically diverse groups. For the majority of multiple regression analyses in this investigation, a regression procedure similar to that of Rego and Bryant (1993) and Näslund and Schneider (1999a, 1999b) was utilized. The hierarchical (sequential) forced entry method was selected for the analyses in this investigation based on two characteristics of the analyses as outlined by Field (2000): 1) the large amount of empirical information available regarding the predictors and 2) the fact that the analyses are used for theory testing.

Each covariate was entered individually with the forced entry method. The nonverbal cognitive abilities measure was always entered as the first variable into the regression equation followed by the other predictor variables proposed by the theoretical model. This was done to ensure that the impact of the remaining predictors on the criterion variable was not confounded with base cognitive skills. For each analysis, the predictors were entered in the same order: cognitive abilities, verbal memory, listening comprehension (where applicable), decoding (where applicable), vocabulary, then phonological awareness. Depending on the purpose of the analysis, group membership and an interaction term were entered last. The predictor(s) in question for the corresponding hypotheses were always entered last. Entering the covariates hierarchically allowed for the careful consideration of the contribution of each predictor separately as well as the changes in R^2 values between each step. Therefore, the adjusted R^2, the change in R^2, the final beta, and the statistical significance of each are reported for every predictor.

As a second step to the hierarchical regressions in which each predictor variable is entered separately into the regression equation, follow-up analyses were conducted to test for significant group differences in the regression slopes that appeared to be different in the bilingual and monolingual groups. In order to show that two samples respond differently to changes in continuous predictors, multiple regressions must be augmented by formally testing the difference in regression coefficients as measured by the group-by-predictor interaction (Aiken & West, 1996; Dallal, 2004). The procedure is described in detail below.

H2a: Word decoding performance will be predicted by similar factors in both groups

Before examining the data with hierarchical regression analyses, correlation coefficients were computed among the hypothesized predictor variables and the word decoding measures at each time of measurement. The results of these descriptive preliminary analyses are presented in Table 34. Using the Bonferroni approach to control for Type 1 error across the 20 correlations, a p-value of less than .0025 (.05/20 = .0025) was required for significance. With this approach, coefficients above $r = .32$ for the Turkish bilingual sample and above $r = .39$ for the German monolingual group were required to meet the stricter significance cutoff. In both groups, the phonological awareness scales were most closely related to the measures of word decoding. With only one exception in the German group, phonological awareness correlated with decoding at the $p < .0025$ level at every time of measurement for both groups. Some meaningful correlations were detected among the measures of vocabulary and decoding, but they are generally weaker than the correlations with the phonological measures. Cognitive abilities seemed to be closely related to decoding at T0 for the monolingual group, and moderately correlated with decoding in the bilingual group. To control for the role of cognitive abilities on decoding, it is further included as a predictor in the following analyses. Verbal memory correlated weakly with decoding and is taken into consideration in the regression analyses below, but because listening comprehension appears to be unrelated to word decoding, it is disregarded for the remaining analyses of decoding. The correlations suggest that of all the base reading skills, phonological awareness is most closely related to word decoding for both groups.

No correlations appeared to differ significantly between the Turkish-German bilingual and German monolingual groups. Overall, the results suggest that the same components are related to word reading in both groups: verbal memory, cognitive abilities, phonological awareness, and German expressive vocabulary skills. These variables will therefore be considered as predictors of word decoding for both groups in the following regression analyses, whereas there was some indication that phonological awareness would emerge as the strongest predictor of word decoding.

Table 34

Correlation Coefficients for all Predictor Variables for Word Decoding at each Time of Measurement for Turkish Bilingual (TB) and German Monolingual (GM) Participants

	TB				GM			
	T0	T1	T2	T3	T0	T1	T2	T3
Cognitive abilities (T -1)	.26*	.24*	.16	.25*	.44**	.28*	.25	.23
Verbal short-term memory (T1)	.24*	.31*	.23*	.32**	.29*	.33*	.29*	.25
Phonological awareness								
T1	.53**	.56**	.51**	.46**	.56**	.47**	.49**	.32*
T2	.44**	.49**	.50**	.36**	.59**	.53**	.60**	.40**
German vocabulary								
T1	.18	.34**	.35**	.32**	.39**	.31*	.24*	.20
T2	.22*	.32**	.32**	.29**	.46**	.34*	.30*	.24
Listening comprehension (T2)	.14	.22*	.20	.20	.18	.13	.21	.02

Note. Using a Bonferroni approach to control for Type 1 errors across the 20 correlations for each group resulted in a p-value cutoff of .0025 to indicate a significant correlation.
*p < .05. **p < .0025 (Bonferroni cutoff).

In total, three hierarchical multiple regressions were conducted for each group to predict word decoding at the end of second grade and in the middle of third grade. In the following analyses, word decoding at T2 and T3 were treated as criterion variables with predictor sets from T1 and T2. The first analysis examined cognitive abilities and the verbal measures from the middle of second grade (T1) as predictors for decoding at the end of second grade (T2). The second analysis used the verbal predictors from T1 to estimate decoding performance one year later at T3, and the final analysis examined the predictive power of the verbal measures at T2 to estimate decoding six months later at T3. By entering each variable separately, it was possible to determine which variables significantly added to the predictive power of the equation. The three sets of regressions are shown together in Table 35 with the adjusted R^2, the change in R^2, and final β weight for each step.

In the first set, the final regression equation with all steps produced a significant R^2 for both the Turkish bilingual group and the German monolingual group, adjusted R^2 = .27, F (4, 81) = 8.31, p <.01 and adjusted R^2 = .18, F (4, 48) = 3.77, p = .01 respectively. Although the significant changes in R^2 seem to indicate that verbal memory and German vocabulary at T1 play a more important role in the bilingual group, phonological awareness remains the only significantly meaningful predictor of word decoding in both groups when all variables are entered into the equation. Overall, several small differences in the change in R^2 for each step, the different beta weights for phonological awareness, and the difference in the final amount of explained variance for each group (27% for the bilingual group compared to 18% for the monolingual group) could be interpreted as indicators of different mechanisms at work in facilitating word reading among the two groups. Comparing separate multiple regression equations does not provide statistically useful evidence regarding group differences. However, the comparisons of these separate equations indicated which predictors could be tested by way of an interaction term, explained in detail below.

The T1 predictors for decoding abilities one year later at T3 provided somewhat different results. Only for the Turkish bilingual group was a significant R^2 found after entering all four steps, adjusted $R^2 = .22$, $F\,(4, 77) = 6.25$, $p < .01$. Again, although all predictors led to a significant change in R^2 for the bilingual group, only phonological awareness at T1 was a significant predictor of decoding at T3. In contrast to the 22% of variance that the four predictors were able to explain for the Turkish bilingual group, the predictors accounted for only 4% of the variance in decoding performance at T3 in the monolingual group. None of the T1 predictors led to a significant change in R^2. The following section describes analyses conducted to examine if this discrepancy stems from a significant difference in the predictive power of phonological awareness between the two groups.

Similarly, when examining the verbal predictors from T2 (in conjunction with cognitive abilities measured at T-1 and verbal memory from T1) in relation to word decoding six months later at T2, a significant R^2 was again found only for the bilingual group, $R^2 = .18$, $F\,(4, 77) = 5.32$, $p < .01$. Among the bilingual group, none of the predictors demonstrated significant predictive power when all four steps were entered into the equation, but the first three steps all produced significant changes in the R^2. Conversely, in the monolingual group, phonological awareness continued to be the only predictor that significantly accounted for variance in decoding performance. Nonetheless, taken together, the four predictors were only able to account for 12% of the variance in decoding at T3 for the monolingual group. The discrepancy in the phonological awareness regression coefficients is examined more closely in the next set of analyses.

Table 35
Separate Hierarchical Multiple Regressions Depicting the Contribution of Cognitive Abilities, Verbal Memory, German Vocabulary, and Phonological Awareness for the Prediction of Word Decoding for the Turkish Bilingual (TB) and German Monolingual (GM) Groups[5]

The relative contribution of T1 verbal predictors for the prediction of word decoding at T2						
Predictors	TB			GM		
	Adjusted R^2	Δ R^2	Final β	Adjusted R^2	Δ R^2	Final β
Step 1						
Cognitive abilities (T-1)	.01	.03	-.03	.04	.06	.06
Step 2						
Verbal memory (T1)	.05	.05*	-.02	.09	.06	.10
Step 3						
German vocabulary T1	.14	.09**	.16	.09	.02	.05
Step 4						
PA at T1	.27	.14**	.48**	.18	.10*	.39*

The relative contribution of T1 verbal predictors for the prediction of word decoding at T3						
Predictors	TB			GM		
	Adjusted R^2	Δ R^2	Final β	Adjusted R^2	Δ R^2	Final β
Step 1						
Cognitive abilities (T-1)	.05	.07*	.09	.03	.05	.19
Step 2						
Verbal memory (T1)	.11	.07*	.10	.06	.06	.18
Step 3						
German vocabulary T1	.16	.06*	.16	.05	.01	.07
Step 4						
PA at T1	.22	.06**	.33**	.04	.01	.12

The relative contribution of T2 verbal predictors for the prediction of word decoding at T3						
Predictors	TB			GM		
	Adjusted R^2	Δ R^2	Final β	Adjusted R^2	Δ R^2	Final β
Step 1						
Cognitive abilities (T-1)	.05	.07*	.12	.03	.05	.12
Step 2						
Verbal memory (T1)	.11	.07*	.21	.06	.05	.11
Step 3						
German vocabulary T2	.15	.05*	.15	.04	.00	-.09
Step 4						
PA at T2	.18	.04	.22	.12	.10*	.39*

Note: PA = Phonological awareness. Significance levels for the change in R^2 refer to the significance in the change in F values for each entered step of the regression. For purposes of simplification, the F values are not displayed here, only R^2.
* $p < .05$. ** $p < .01$.

[5] In this and all following multiple regression tables, Final β refers to the standardized β coefficient when all steps are entered into the equation.

The regressions presented here do not provide conclusive support for Hypothesis 2a, but they are not contradictory either. In all six regression equations, phonological awareness was the only variable that reached significance as a coefficient in the final regression equation. Since phonological awareness was the most important predictor both groups, there is reason to believe that the same predictors are chiefly responsible for facilitating word decoding in both groups and that H2a (*Word decoding performance will be predicted by similar factors in both groups)* was initially supported.

To follow up on the discrepancies found above with regard to the power of phonological awareness in predicting word decoding performance, an additional set of stepwise multiple regressions included an interaction term to test for significant group differences. For this investigation, the primary purpose of the separate regression models was to view the regression equations of each group separately, and then to select the predictors that appeared divergent for the two groups for testing with an interaction term. Simply fitting separate regression models and declaring them to differ based only on significance levels is not adequate for determining statistically significant group differences. (e.g., Dallal, 2004). Although this procedure is essentially identical to the individual regression equations that are obtained by fitting the linear regression of the base reading variables for the monolingual and bilingual participants, this approach allows for testing whether the regression coefficients differ between the two groups. A significant interaction term regression coefficient indicates significant differences in the slopes of two groups.

For the final step in investigating H2a, the three sets of hierarchical multiple regressions above were conducted again, but with both groups simultaneously. This required several modifications to the previous analyses. First, the variable for which differences were being tested (in this case, phonological awareness) was transformed into a z-value to reduce the possibility of confounding with the interaction term. Second, a dummy variable denoting the two groups was integrated into the equation (1 = monolingual, 0 = bilingual). Finally, a last variable representing the interaction between the variable of interest and the group was created, transformed into a z-value, and entered into the equation as an interaction term. If the interaction term is significant in a multiple regression analysis of this nature, it can be assumed that the groups differ significantly on the variable in question with regard to its power in predicting the criterion variable. The three combined-sample regression analyses are presented together in one table, Table 36.

Table 36

Stepwise Hierarchical Multiple Regressions with Verbal Predictors and Interaction Terms for Word Decoding for all Participants

The relative contribution of verbal predictors from T1 for the prediction of word decoding at T2			
Predictors	R^2	Δ R^2	Final β
Step 1	.03	.04*	
Cognitive abilities (T-1)			.00
Step 2	.07	.05**	
Verbal memory (T1)			.03
Step 3	.14	.07**	
German vocabulary T1			.15
Step 4	.25	.12**	
PA at T1			.43**
Step 5	.25	.00	
Language group (1= Bilingual, 0 = Monolingual)			-.11
Interaction term: PA x language group			.05
The relative contribution of verbal predictors from T1 for the prediction of word decoding at T3			
Predictors	R^2	Δ R^2	Final β
Step 1	.06	.06**	
Cognitive abilities (T-1)			.10
Step 2	.10	.05**	
Verbal memory (T1)			.13
Step 3	.15	.05**	
German vocabulary T1			.16
Step 4	.18	.04*	
PA at T1			.14
Step 5	.17	.01	
Language group (1= Bilingual, 0 = Monolingual)			-.23
Interaction term: PA x language group			.22
The relative contribution of verbal predictors from T2 for the prediction of word decoding at T3			
Predictors	R^2	Δ R^2	Final β
Step 1	.06	.07**	
Cognitive abilities (T-1)			.12
Step 2	.10	.05*	
Verbal memory (T1)			.16
Step 3	.13	.04*	
German vocabulary T2			.07
Step 4	.17	.05*	
PA at T2			.29*
Step 5	.17	.01	
Language group (1= Bilingual, 0 = Monolingual)			-.08
Interaction term: PA x language group			-.05

Note: PA = Phonological awareness. Significance levels for the change in R^2 refer to the significance in the change in F values for each entered step of the regression. For purposes of simplification, the F values are not displayed here, only R^2. Both the interaction term and the relevant variable (here PA) were z-transformed to decrease confoundability.
* $p < .05$. ** $p < .01$.

All three equations produced viable solutions with significant results (T1/T2: $R^2 = .25$, F (6, 132) = 8.15, $p < .01$; T1/T3: $R^2 = .17$, F (6, 118) = 5.07, $p < .01$; T2/T3: $R^2 = .17$, F (6, 117) = 5.19, $p < .01$). In the first equation with predictors from the middle of second grade and the criterion variable from the end of second grade, only phonological awareness demonstrated a significant regression coefficient in predicting decoding. Although no T1 variables were able to significantly predict decoding skills a year later at T3, only phonological awareness emerged again in the third analysis as the only T2 variable able to significantly predict decoding in the third grade.

One final follow-up regression, not displayed in Table 36, was conducted to investigate the possible significant interaction for German vocabulary skills, hinted at in the T2 to T3 separate group analysis, in which vocabulary was found to make a significant R^2 contribution to predicting decoding skills in the bilingual group but none at all in the monolingual group. The German vocabulary interaction term in this equation played no significant role in the final regression equation, while the remaining predictors did not change substantially, $R^2 = .22$, F (6, 117) = 5.41, $p < .01$.

Overall, the hierarchical regression models demonstrated no significant differences in the predictive strength of cognitive abilities, verbal memory, phonological awareness, or German vocabulary variables between the German monolingual and Turkish bilingual groups for the prediction of decoding abilities. It should, however, be kept in mind that only between 18 and 27 percent of the variance in decoding performance was accounted for in the Turkish bilingual group and only between 4 and 18 percent in the German monolingual group. Therefore, although it seems that, with regard to the predictive power of the variables at hand, the groups did not differ, there is still around 80% of the variance in decoding abilities left unexplained. It is possible that unknown opaque predictors play varied roles amongst the two groups. Until that is confirmed however, it will be assumed that decoding is predicted similarly in both groups and that Hypothesis 2a is plausible.

H2b: Reading comprehension performance will be predicted by the same core components of reading abilities for both the bilingual and monolingual groups

Before examining the predictors of reading comprehension, it is helpful to provide the mean scores and standard deviations for the reading comprehension measures at the end of second grade (T2) and the middle of third grade (T3). Although there was a slight tendency for the German monolingual group to perform better, as shown in Table 37, the groups did not differ meaningfully in their performance on the reading comprehension scale at either T2 or T3, $t(159) = -1.23$, $p = .22$ and $t(144) = -1.37$, $p = .17$. Both groups improved significantly in their performance over the six months, and no difference in the amount of growth in reading skills over those six months was found, $t(138) = -.71$, $p = .48$[6]. Nonetheless, the lack of mean differences does not provide any indication of potential structural differences among the factors facilitating reading comprehension. The next sets of analyses investigate the factors responsible for reading comprehension.

Table 37

Means and Standard Deviations for the Reading Comprehension Measures at the End of 2nd Grade (T2) and at the Middle of 3rd Grade (T3) for the Turkish Bilingual and German Monolingual Groups

	TB	GM
Reading comprehension T2		
M	6.13	6.75
SD	2.92	3.44
Reading comprehension T3		
M	8.35	9.28
SD	6.31	4.43

A procedure similar to that utilized for H2a was followed to examine the core predictors of reading comprehension among the Turkish bilingual and German monolingual children in the second and third grades. A series of Pearson product-moment correlations were first calculated for both groups separately to look for significant differences in the relationships between the predictor variables proposed by Näslund and Schneider (1991) and the reading comprehension measures. According to H2b, it was expected that the proposed predictors would be related to reading for both groups.

Table 38 shows the correlations between the proposed predictors and reading comprehension for both groups. The Bonferroni approach to controlling for Type 1 errors across the 18 correlation coefficients would call for a *p*-value cut-off of .003. For the Turkish bilingual group, reading comprehension at T2 and T3 correlated significantly at the Bonforroni cut-off level with the measures of phonological awareness at both T1 and T2, the German vocabulary measures at both times of measurement, and the decoding measures at both T1 and T2. The same was true for the German monolingual group with the exception of the correlations between German vocabulary at T1 and T2 with reading comprehension at T2, in

[6] Development of reading comprehension skills are discussed in detail in Section 4.4..

which the correlations were significant at the .05 and .01 levels, but not under the Bonferroni cutoff of .003.

Cognitive abilities appeared only weakly related to reading comprehension, but will remain in the following analyses as a control variable[7]. Verbal memory was moderately significantly related to reading comprehension in the third grade and will also be maintained as a predictor variable in the next set of analyses. Furthermore, as key component of the Näslund and Schneider (1991) model, verbal memory is essential for further testing of the model. Listening comprehension was (moderately) related to reading comprehension for the bilingual group only. It will also remain in the following regression analyses for testing potential differential predictive power. Nonetheless, due to its uncertain psychometric properties, the listening comprehension scale is interpreted with caution.

Table 38

Correlation Coefficients of all Possible Predictor Variables for Reading Comprehension at each Time of Measurement for Turkish Bilingual (TB) and German Monolingual (GM) Participants

		TB		GM	
		T2	T3	T2	T3
Cognitive abilities (T -1)		.15	.20	.19	.33*
Verbal short-term memory (T1)		.20	.23*	.20	.31*
Phonological awareness					
	T1	.49**	.55**	.47**	.57**
	T2	.39**	**.38****	.56**	**.71****
German vocabulary					
	T1	.47**	.36**	.26*	.41**
	T2	.38**	.34**	.35*	.48**
Listening comprehension (T2)		.32**	.35**	.19	.14
Decoding					
	T1	.54**	.66**	.53**	.65**
	T2	.59**	.59**	.72**	.76**

Note. Using a Bonferroni approach to control for Type 1 errors across the 18 correlations for each group resulted in a *p*-value cutoff of .003 to indicate a significant correlation. Coefficients that significantly differed between the German monolingual and Turkish bilingual groups have been marked bold.
**p* < .05. ***p* < .003 (Bonferroni cutoff).

[7] The decision to maintain cognitive abilities as a control variable in the following analyses was based on the majority of other studies examining models of reading development (e.g., Comeau et al., 1999; Schneider, 2004) and the recommendation of Wagner and collaborators (1987) to control for cognitive abilities in order to avoid the possibility of overestimating the actual relationship between phonological awareness and reading.

One significant difference emerged with regard to the correlation coefficients in the two groups. Phonological awareness measured at the end of second grade demonstrated a much stronger correlation with reading comprehension in third grade in the monolingual group (r = .71) than in the bilingual group (r = .38), t (144) = 3.05, $p < .01$. This will be explored in greater detail in the regression analyses below. Nonetheless, with the exception of listening comprehension, the same variables demonstrated significant correlations with reading comprehension in both groups. Therefore, H2b was supported in part by the data. The regression analyses to follow will provide additional information for the hypothesis that the same core abilities are related to reading comprehension in both groups.

H2c & H2d: Phonological awareness will be a weaker predictor of reading comprehension among Turkish-German bilingual children as compared to monolingual German children (H2c); in turn, vocabulary will be a stronger predictor of reading comprehension abilities for the Turkish-German bilingual readers than for the German monolingual readers (H2d)

A series of hierarchical stepwise regressions identical to those calculated to predict word decoding were computed to explore the predictive powers of five variables in estimating reading comprehension abilities at the end of second grade and the beginning of third grade. The five variables proposed in the theoretical model of reading (verbal memory, decoding, German vocabulary, and phonological awareness) were tested as predictors of reading comprehension along with the measure of cognitive ability, included as a control. Also parallel to the analyses of decoding, predictors from T1 were used to estimate reading comprehension at T2, followed by analyses of predictors from T1 to estimate reading comprehension at T3, and finally, predictors from T2 were used to predict reading at T3. Since listening comprehension was only measured at T2, it is only included as a predictor for reading at T3 in the third set of analyses. Each analysis was performed separately for the two groups to gain a sense of differential regression coefficients within each equation. Table 39 provides a summary of all six analyses.

The linear combinations of the predictor variables from each time of measurement were significantly related to reading comprehension at the end of second grade (T2) (German monolingual predictors at T1: F (5,46) = 15.09; Turkish bilingual predictors at T1: F (5,77) = 15.19) and in the middle of third grade (T3) (German monolingual T1 predictors: F (5,37) = 10.30, T2 predictors: F(6,34) = 16.44; Turkish bilingual predictors at T1 F (5,73) = 11.70, T2 predictors: F (6,71) = 8.84) for both groups at a $p < .001$ level. Based on these results, it is clear that the independent variables are good predictors of reading comprehension at both T2 and T3. In the first set of analyses, decoding and phonological awareness at T1 appear to be the strongest predictors of reading comprehension at T2, with a total explained variance across all five predictors of 47% for the Turkish bilingual group and 58% for the German monolingual group. In neither group did German vocabulary play a meaningful role when entered into the T1 to T2 equation.

An almost identical final equation resulted from the regression of T3 reading comprehension onto the predictor variables from one year earlier at T1. Vocabulary added no additional predictive power overall, but it did provide a significant change in the R^2 for the bilingual group when entered. Also, both cognitive abilities and verbal memory provided significant additional predictive power when entered into the monolingual equation, but not in the bilingual equation. However, decoding and phonological awareness again proved to be the most important predictors of reading in third grade for both groups. As with the previous set, the analyses explained more variance for in the monolingual group than for the bilingual group (53% compared to 42%).

The set of predictors from T2 showed more differential regression coefficients for the two groups in predicting reading comprehension a half year later at T3. Listening comprehension produced a significant change in the R^2 upon entry for the bilingual group but not the monolingual group. In the final regression equation however, only one predictor, decoding, significantly accounted for variance in reading comprehension in the bilingual group, for a total explained variance of 38%. Within the monolingual group, on the other hand, both decoding and phonological awareness were strong predictors of reading comprehension in the third grade with a total explained variance across all five predictors of 70%.

These analyses uncovered several discrepancies between the Turkish bilingual and German monolingual groups in the prediction of reading comprehension. In light of H2c, phonological awareness did in fact seem to play a much weaker role in the prediction of reading comprehension among the bilingual children, particularly in the third set of predictors (from T2) in estimating reading performance in third grade (bilingual $\beta = .08, p < .48$; monolingual $\beta = .43, p < .01$). In order to test this difference for significance, the next set of analyses provides a further series of multiple regressions with interaction terms.

With regard to H2d, there was little indication that German vocabulary played a stronger predictive role in the reading comprehension equations for the bilingual group. Although it did significantly add to the predictive power of the equation from T1 to T3 for the bilingual group only, the opposite was true within the third set of analyses from T2 to T3. Since there appeared to be no empirical indication to justify further testing for group differences in German vocabulary skills as a predictor of reading comprehension, only differences in phonological awareness predictive powers will be tested in the next series of multiple regressions.

Table 39

Separate Hierarchical Multiple Regressions Depicting the Contribution of Cognitive Abilities, Verbal Memory, Listening Comprehension, Decoding, German Vocabulary, and Phonological Awareness to Reading Comprehension for the Turkish Bilingual (TB) and German Monolingual (GM) Groups

The relative contribution of T1 predictors for the prediction of reading comprehension at T2

Predictors	TB			GM		
	Adjusted R^2	Δ R^2	Final β	Adjusted R^2	Δ R^2	Final β
Step 1						
Cognitive abilities (T-1)	.01	.02	-.10	.02	.03	-.13
Step 2						
Verbal memory (T1)	.04	.04	-.04	.03	.04	-.17
Step 3						
Decoding T1	.43	.39**	.55**	.54	.50**	.72**
Step 4						
German vocabulary T1	.45	.02	.10	.53	.00	.02
Step 5						
PA at T1	.47	.03*	.23*	.58	.05*	.29*

The relative contribution of T1 predictors for the prediction of reading comprehension at T3

Predictors	TB			GM		
	Adjusted R^2	Δ R^2	Final β	Adjusted R^2	Δ R^2	Final β
Step 1						
Cognitive abilities (T-1)	.02	.04	-.06	.09	.11*	.02
Step 2						
Verbal memory (T1)	.04	.03	-.11	.17	.10*	.02
Step 3						
Decoding T1	.27	.24**	.34**	.43	.26**	.51**
Step 4						
German vocabulary T1	.33	.07**	.18	.45	.03	.10
Step 5						
PA at T1	.42	.09**	.41**	.53	.08*	.33*

The relative contribution of T2 predictors for the prediction of w reading comprehension at T3

Predictors	TB			GM		
	Adjusted R^2	Δ R^2	Final β	Adjusted R^2	Δ R^2	Final β
Step 1						
Cognitive abilities (T-1)	.02	.03	.03	.09	.12*	-.02
Step 2						
Verbal memory (T1)	.03	.03	-.03	.18	.10*	.01
Step 3						
Listening comp. (T2)	.08	.07*	.15	.16	.00	-.11
Step 4						
Decoding T2	.38	.30**	.52**	.53	.36**	.49**
Step 5						
German vocabulary T2	.39	.01	.10	.59	.07*	.20
Step 6						
PA at T2	.38	.00	.08	.70	.11**	.43**

Note: PA = Phonological awareness. Significance levels for the change in R^2 refer to the significance in the change in *F* values for each entered step of the regression. For purposes of simplification, the *F* values are not displayed here, only R^2.
* $p < .05$. ** $p < .01$.

Table 40 provides an overview of the stepwise hierarchical multiple linear regressions that were calculated with interaction terms to test the extent to which Hypothesis 2c could be supported with significance tests. Again, an interaction term with group membership and phonological awareness abilities was built into a collapsed version of the regression equation with both groups together to test the hypothesis that phonological awareness plays a significantly weaker role in reading comprehension for Turkish bilingual children. A summary of the analyses for each combination of predictor measurement points and criterion measurement points (T1 to T2, T1 to T3, and T2 to T3) is shown in Table 40.

The linear combination of all predictors was significantly related to reading comprehension in each set of analyses at a significance level of $p < .001$ (T1 predictors of T2 reading comprehension: F (7, 127) = 21.69, T1 predictors of T3 reading comprehension: F(7,114) = 15.94, T2 predictors of T3 reading comprehension: F(8,110) = 17.94). In the first set, in which predictors from the middle of second grade (T1) were used to estimate reading comprehension at the end of second grade (T2), two variables were found to be statistically significant predictors in the final equation: decoding and phonological awareness. Neither group membership nor the interaction term produced meaningful effects, thus indicating that the set of predictors, including phonological awareness predicted reading comprehension for both groups in similar ways. The second analyses, in which reading comprehension at T3 was regressed onto predictors from T1, produced very similar results. Decoding and phonological awareness were the only significant predictors of reading in the third grade in the final regression equation. Again, no indication of group differences was found in the regression coefficients of the group membership variable or the interaction term.

The third analysis, examining the predictive influence of the variables from T2 on reading in third grade, produced a significant interaction effect. Again, phonological awareness and decoding were significant predictors, but the largest regression coefficients were found for the group membership dummy variable and the interaction term for phonological awareness and group membership. The strong negative regression coefficient indicated a significantly stronger effect for the group designated with “0” (the monolingual group) for the phonological awareness variable.

Table 40

Hierarchical Multiple Regressions with all Predictors and Interaction Terms for Reading Comprehension for all Participants

The relative contribution of T1 predictors for the prediction of reading comprehension at T2			
Predictors	Adjusted R^2	ΔR^2	Final β
Step 1	.02	.03	
Cognitive abilities			-.10
Step 2	.04	.03*	
Verbal memory			-.09
Step 3	.48	.44**	
Decoding			.62**
Step 4	.49	.02*	
German vocabulary T1			.07
Step 5	.52	.03**	
PA at T1			.33**
Step 6	.52	.01	
Language group (1= Bilingual, 0 = Monolingual)			.09
Interaction term: PA x language group			-.18
The relative contribution of T1 predictors for the prediction of reading comprehension at T3			
Predictors	Adjusted R^2	ΔR^2	Final β
Step 1	.05	.06**	
Cognitive abilities			-.02
Step 2	.09	.04*	
Verbal memory			-.04
Step 3	.35	.26**	
Decoding			.41**
Step 4	.39	.05**	
German vocabulary T1			.19*
Step 5	.47	.07**	
PA at T1			.45**
Step 6	.46	.01	
Language group (1= Bilingual, 0 = Monolingual)			.19
Interaction term: PA x language group			-.26
The relative contribution of T2 predictors for the prediction of reading comprehension at T3			
Predictors	Adjusted R^2	ΔR^2	Final β
Step 1	.05	.06**	
Cognitive abilities			.01
Step 2	.08	.04*	
Verbal memory			.00
Step 3	.11	.04*	
Listening comprehension			.04
Step 4	.45	.34**	
Decoding			.51**
Step 5	.48	.03*	
German vocabulary T2			.16
Step 6	.50	.03**	
PA at T2			.45**
Step 7	.53	.04*	
Language group (1= Bilingual, 0 = Monolingual)			.55**
Interaction term: PA x language group			-.62**

Note: PA = Phonological awareness. Significance levels for the change in R^2 refer to the significance in the change in F values for each entered step of the regression. For purposes of simplification, the F values are not displayed here, only R^2. Both the interaction term and the relevant variable (here PA) were z-transformed to decrease confoundability
* $p < .05$. ** $p < .01$.

Taken together, these regression analyses provide some support for Hypothesis 2c. In one of the three analyses with built-in interaction terms, a significant interaction between phonological awareness and group membership was detected. It could therefore be assumed that when predicting reading for bilingual and monolingual children from the second to third grades, phonological awareness was a more important component of reading comprehension for the monolingual children than it was for the bilingual children. Hypothesis 2d predicted that German vocabulary would be a stronger predictor for reading comprehension among bilingual children. This hypothesis was not supported by the analyses[8].

4.4. Patterns of development

This section aims to address questions regarding possible differential patterns of literacy development among bilingual and monolingual children in early primary school. An overview of intercorrelations is first provided to show the relationships among gain scores for all aspects of verbal abilities and literacy that were measured longitudinally in this study. Next, correlates of growth are examined for both word reading and reading comprehension. Finally, by way of multiple regressions, predictors of growth and differences in those predictors are examined.

Before investigating the factors contributing to the patterns of literacy development in the two groups, analyses were conducted to examine the extent to which the development in verbal and literacy skills correlate, possibly indicating a general factor of growth underlying all aspects of verbal and literacy development. In addition, cognitive abilities were correlated with the gain scores to see if cognitive skills are associated with the rate of growth. Because word decoding was measured multiple times, three separate gain scores[9] were calculated and used in the analyses. The Pearson product-moment correlation coefficients were computed for all gain scores and are shown in Table 41 for the Turkish bilingual group and in Table 42 for the German monolingual group.

[8] In order to control for the possibility that the strong effects of decoding detract from the effects of phonological awareness, all regression analyses in for H2c and H2d were conducted without the decoding variable. The results were essentially the same. An interaction effect for phonological awareness by group was found only in the T2 to T3 analyses, only with a lower amount of explained variance. These analyses can be found in Appendix C, Tables C5 and C6.

[9] *Gain scores* refer to simple change scores or difference scores calculated by subtracting the pre-test from post-test scores (Becker, 2000). For a review of gain scores in longitudinal research of this nature see Hendrickson and Jones (1987).

Table 41

Correlations between Gain Scores for all Longitudinal Measures for the Turkish Bilingual Group

	Cognitive abilities	PA (T1 to T2)	GV[a] (T1 to T2)	Decoding (T1 to T2)	Decoding (T2 to T3)	Decoding (T0 to T3)
PA (T1 to T2)	-.10	--				
GV[a] (T1 to T2)	-.11	.13	--			
Decoding (T1 to T2)	-.15	.10	-.03	--		
Decoding (T2 to T3)	**.21**	-.10	-.01	-.29*	--	
Decoding (T0 to T3)	.06	-.05	-.05	.37**	.40**	--
Reading comp. (T2 to T3)	.08	-.13	-.18	.23*	.09	.07

Note. PA = Phonological awareness, GV = German vocabulary. Using a Bonferroni approach to control for Type 1 errors across the 36 correlations for each group resulted in a *p*-value cutoff of .001 to indicate a significant correlation. Coefficients that significantly differed between the German monolingual and Turkish bilingual groups have been marked bold.

[a] The German vocabulary gain score is based on the whole scale raw scores on picture vocabulary and synonyms at T1 and T2

* $p < .05$. ** $p < .001$ (Bonferroni cutoff).

Table 42

Correlations between Gain Scores for all Longitudinal Measures for the German Monolingual Group

	Cognitive abilities	PA (T1 to T2)	GV[a] (T1 to T2)	Decoding (T0 to T3)	Decoding (T1 to T2)	Decoding (T2 to T3)
PA (T1 to T2)	-.19	--				
GV[a] (T1 to T2)	.09	.28*	--			
Decoding (T1 to T2)	-.02	.06	.04	--		
Decoding (T2 to T3)	**-.17**	.07	.03	-.36*	--	
Decoding (T0 to T3)	-.13	.19	-.12	.32*	.56**	--
Reading comp. (T2 to T3)	.19	.18	-.04	.15	.26	.32*

Note. PA = Phonological awareness, GV = German vocabulary. Using a Bonferroni approach to control for Type 1 errors across the 36 correlations for each group resulted in a *p*-value cutoff of .001 to indicate a significant correlation. Coefficients that significantly differed between the German monolingual and Turkish bilingual groups have been marked bold.

[a] The German vocabulary gain score is based on the whole scale raw scores on picture vocabulary and synonyms at T1 and T2

* $p < .05$. ** $p < .001$ (Bonferroni cutoff).

Among the gain scores for the Turkish bilingual group, no (highly) significant correlations were found between measures. Only the decoding gain scores were moderately related. In the German monolingual group, there was some evidence that growth in phonological awareness was related to growth in vocabulary skills. A moderate correlation was also found between the decoding gain score from the end of second grade to the middle of third grade (T2 to T3) and growth in reading comprehension performance. But also within the monolingual group, only the correlations between gain scores for decoding at the different time intervals reached a high level of significance.

The correlation coefficients were also tested for significant differences between the two groups with the Fisher r to z transformation. One significant difference was detected. In the bilingual group, cognitive abilities seemed to be positively related to the gain score in word decoding from the middle of second grade to mid-third grade ($r = .21$), while this correlation was significantly different and negative for the monolingual group ($r = -.17$; $t\,(128) = 2.41, p < .05$). However, since neither of those correlation coefficients was significant, this discrepancy can be largely disregarded. Overall, the gain scores did not appear to be related to one another, seemingly indicating independent growth processes for each ability in question.

H3a: On measures of phonological awareness, vocabulary, word decoding, and reading comprehension, both the bilingual and monolingual groups will show similar development over time

To investigate the development trends for German vocabulary, phonological awareness, and word decoding within the two groups, as well as possible differences in their rates of development, mixed-design repeated-measure one-way ANCOVAS were conducted for each measure. For each ANCOVA, the dependent variables were the longitudinal measures (phonological awareness, German vocabulary, or word decoding) at each time of measurement and the factor used was time of measurement. Additionally, a between-subject factor, group, was integrated to test for group differences and interactions between group and time. In all analyses, cognitive abilities measured in the middle of first grade and gender were entered into the equation as covariates. Since means and standard deviations have already been presented in Sections 4.2. and 4.3. (Tables 31, 32, and 33), the analyses are depicted separately for each measure graphically in Figures 6, 7, and 8.

The results of the ANCOVAs were relatively uniform across the three scales; several main effects for time and group were clear, but no interaction effects were found. For the phonological awareness scales, when controlling for cognitive skills and gender, there was no significant effect for time, Wilks $\Lambda = .99$, $F\,(1, 143) = 1.08, p = .30$, multivariate $\eta^2 = .01$, and as can be seen in Figure 6, no meaningful group differences were found either, $F(1, 143) = 3.26, p =$

.07, multivariate $\eta^2 = .02$. Overall, the prediction of Hypothesis 3a that the two groups would not differ in phonological awareness development was supported (in fact, no development was detected).

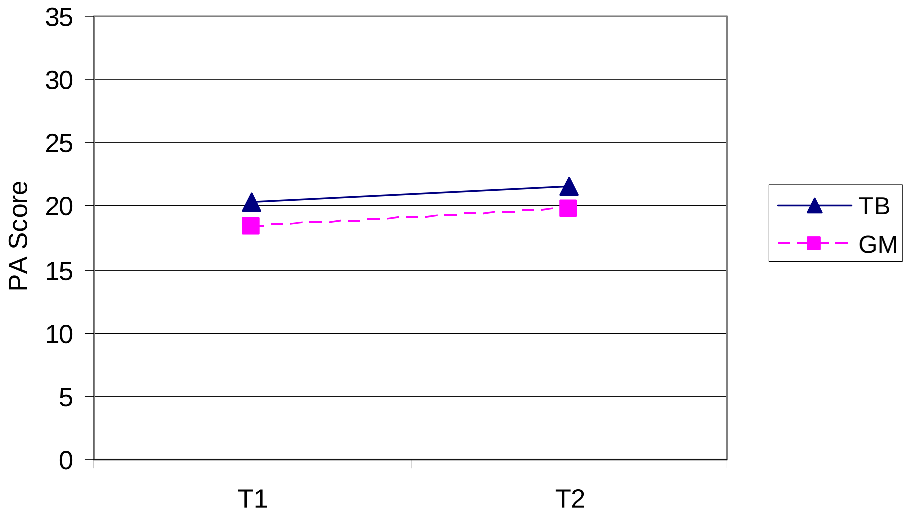

Figure 6. Growth comparisons between the Turkish bilingual (TB) and German monolingual (GM) groups for phonological awareness

The repeated measures analysis of covariance for German vocabulary performance was based on the full item raw scores of the two subscales that were presented at both times of measurement (picture vocabulary and synonyms), displayed in Figure 7. Controlling for cognitive skills and gender, the analyses produced significant effects for both time, Wilks $\Lambda = .91$, F (1, 142) = 13.76, $p < .01$, multivariate $\eta^2 = .09$, and group, F (1, 142) = 49.41, $p < .01$, multivariate $\eta^2 = .26$. These results indicate that regardless of cognitive abilities or gender, both groups improved significantly in their vocabulary skills from the middle to the end of the second grade. This repeated measures ANCOVA also clearly indicated that there were significant differences in German vocabulary skills between the German monolingual and the Turkish bilingual groups. Again, however, no interaction effects were detected. This analysis supported the hypothesis that, in terms of the development rate of German vocabulary, the Turkish bilingual children do not differ from the German monolingual children.

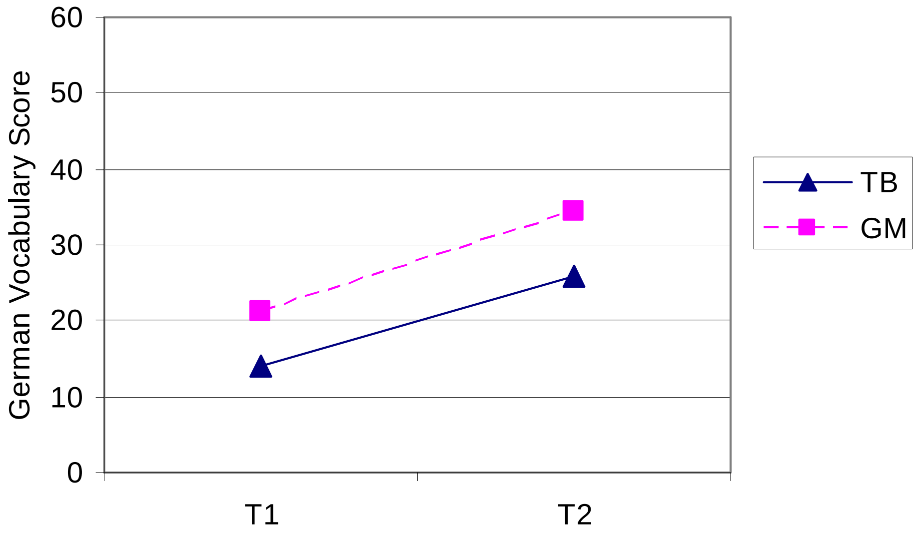

Figure 7. Growth comparisons between the Turkish bilingual (TB) and German monolingual (GM) groups for German vocabulary

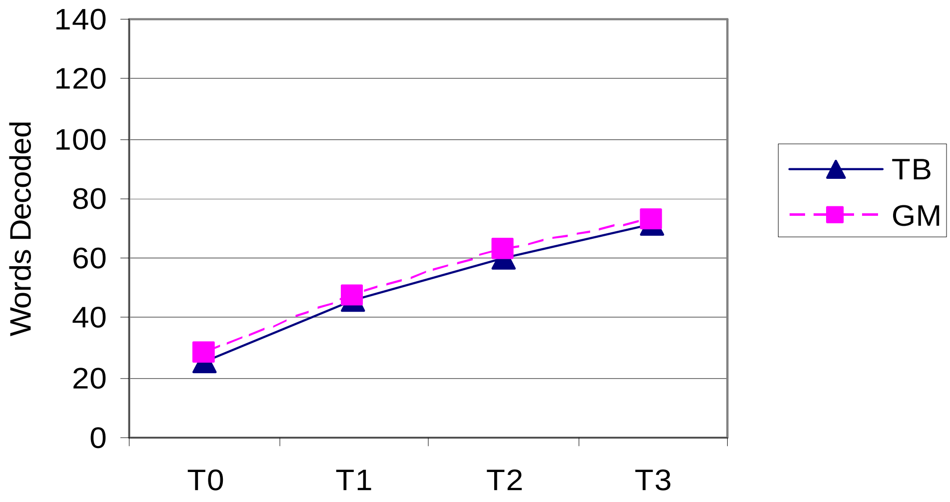

Figure 8. Growth comparisons between the Turkish bilingual (TB) and German monolingual (GM) groups for decoding

With regard to word decoding abilities, Section 4.2. reported that a significant effect for time was found when analyzing the number of words decoded correctly from the end of first grade until the middle of third grade (effect size = .28). Follow-up polynomial contrasts showed a significant linear effect with means increasing significantly over time, F (1,110) = 40.39, $p < .01$, partial $\eta^2 = .27$. The higher-order polynomial contrasts indicated a slight quadratic tendency, however, F (1,110) = 4.05, $p = .05$, partial $\eta^2 = .04$. This finding indicates that both groups improve their decoding skills significantly over time, but that the gains are greater for the earlier measurement points and diminish slightly over time. Not surprising when viewing Figure 8, no group differences could be detected across the four points of measurement, $F(1, 110) = 24$, , $p = .63$, multivariate $\eta^2 = .00$. Again, there were no interaction effects.

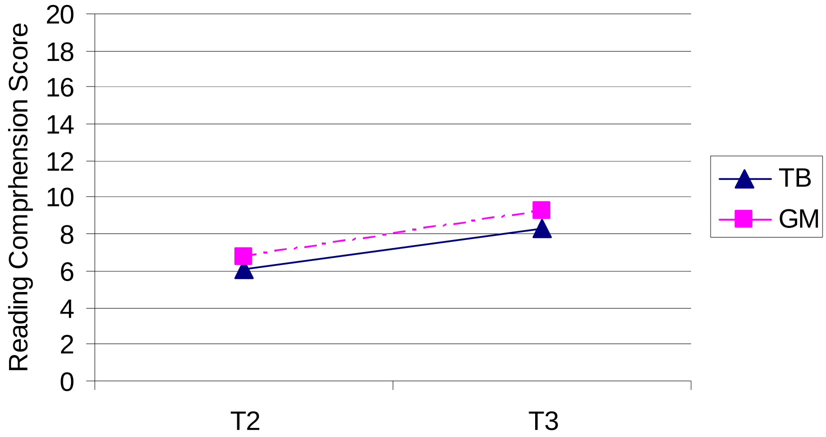

Figure 9. Growth comparisons between the Turkish bilingual (TB) and German monolingual (GM) groups for reading comprehension

Figure 9 visually illustrates the rate of growth in reading comprehension performance over a six-month period from the end of second grade through the middle of third grade. As with the growth analyses above, a repeated measures one-way ANCOVA with *group* as the between-subjects factor was calculated with two covariates: gender and cognitive abilities. Under the constraints of the covariates, no main effect for time was detected, Wilks $\Lambda = 1.00$, $F\,(1, 116) = .31$, $p = .58$, multivariate $\eta^2 = .00$. Similar results were found without the consideration of coavariates. No group differences were detected either, $F\,(1, 116) = 1.50$, $p = .22$, $\eta^2 = .01$. Most importantly, the ANCOVA revealed no significant interactions between time and group. Taken together, the results from these analyses provide the final piece of information required to support Hypothesis H3a: No evidence indicated that Turkish bilingual and German monolingual children differ with regard to their development in phonological awareness, vocabulary development, word decoding, or reading comprehension in their early primary school years.

H3b: The factors predicting growth in word decoding will be similar for both the monolingual and the bilingual groups

Two stepwise hierarchical multiple regression models, summarized in Table 43, were calculated for each group to explore the factors related to growth in word decoding and to investigate possible group differences. For both analyses, the criterion was word decoding in the middle of third grade (T3). In the first model, the regression was conducted with predictors from the middle of second grade (T2). The first step in the regression equation was the entry of word decoding performance at T1. The verbal predictors (German vocabulary and phonological awareness) were then entered separately to see if any additional predictive power could be gained.

For the first model, the final equation was significant for both groups, bilingual $R^2 = .48$, $F(3, 84) = 27.60$, $p < .01$, monolingual $R^2 = .47$, $F(3, 49) = 16.39$, $p < .01$. All of the variance, however, was accounted for by decoding in the middle of second grade (T2). Neither phonological awareness nor German vocabulary contributed any additional predictive power to the equation.

The second model used predictors from the end of second grade (T2) to estimate word reading performance six months later in third grade (T3). Again, the equations were highly significant for both groups, bilingual $R^2 = .69$, $F(3, 82) = 59.84$, $p < .01$, monolingual $R^2 = .71$, $F(3, 47) = 39.75$, $p < .01$, no variable other than the decoding variable at T2 explained the total variance. In other words, no growth in decoding abilities could be accounted for by phonological awareness or German vocabulary performance in either group.

Since the regression weights for both models were nearly identical in each group, it can be assumed that there were no differences between German monolingual children and Turkish bilingual children in the predictive power of either phonological awareness or vocabulary skills when predicting growth in word decoding. It is, however, possible that differences would be found in variables not tested or available for this study. Hypothesis 3b was supported by these analyses: Growth in word decoding was predicted similarly for both German monolingual and Turkish bilingual children.

Table 43

Separate Hierarchical Multiple Regressions Showing the Contribution of German Vocabulary, and Phonological Awareness to Growth in Word Decoding Performance for the Turkish Bilingual (TB) and German Monolingual (GM) Groups

Model	Step	TB			GM		
		Adjusted R^2	ΔR^2	Final β	Adjusted R^2	ΔR^2	Final β
T1 predictors	1 Word decoding	.48	.48**	.62**	.48	.49**	.68**
	2 German vocabulary	.47	.00	.04	.47	.00	-.05
	3 PA	.48	.01	.13	.47	.01	.10
T2 predictors	1 Word decoding	.67	.68**	.80**	.71	.72**	.84**
	2 German vocabulary	.68	.01	.10	.71	.00	.02
	3 PA	.68	.00	-.02	.70	.00	-.01

Note. PA = Phonological awareness. Dependent variable was word decoding at T3. Significance levels for the change in R^2 refer to the significance in the change in F values for each entered step of the regression. For purposes of simplification, the F values are not displayed here, only R^2 values.
* $p < .05$. ** $p < .01$.

H3c & H3d: Phonological awareness will be a weaker predictor of gains in reading comprehension for the Turkish-German bilingual group than for the German monolingual group (H3c), instead, vocabulary skills will be a stronger predictor of gains in reading comprehension for the Turkish-German bilingual group than for the German monolingual group (H3d)

In a procedure similar to the analysis of predictors for growth in decoding abilities, hierarchical multiple regressions were used to test the variables related to growth in reading comprehension abilities. In Table 44 presents a summary of the separately calculated regressions for each group. Reading comprehension in the third grade (T3) is the criterion variable. Reading comprehension at the end of second grade (T2) was the first predictor entered into the regression equation as the first step. The remaining variance was left to be accounted for by phonological awareness, German vocabulary, and word decoding variables at both T1 and T2. The final linear combination of all predictors was highly significant for both groups, bilingual $R^2 = .64$, F (7, 76) = 22.44, $p < .01$, monolingual $R^2 = .69$, F (7, 43) = 17.20, $p < .01$.

Table 44

Separate Hierarchical Multiple Regressions Showing the Contribution of German Vocabulary, Phonological Awareness, and Word Decoding to Growth in Reading Comprehension Performance for the Turkish Bilingual (TB) and German Monolingual (GM) Groups

Step	TB			GM		
	Adjusted R^2	ΔR^2	Final β	Adjusted R^2	ΔR^2	Final β
1 Reading comprehension T2	.59	.59**	.59**	.53	.54**	.31*
2 Decoding T1	.59	.01	-.12	.53	.01	-.09
3 German vocabulary T1	.61	.03*	.21*	.58	.05*	.16
4 PA T1	.63	.02	.17	.60	.03[a]	-.02
5 Decoding T2	.65	.02*	.24*	.64	.05*	.32*
6 German vocabulary T2	.65	.01	-.09	.65	.01	.02
7 PA T2	.64	.00	-.05	.69	.05**	.37**

Note. PA = Phonological awareness. Dependent variable was reading comprehension at T3. Significance levels for the change in R^2 refer to the significance in the change in F values for each entered step of the regression. For purposes of simplification, the F values are not displayed here, only R^2 values.
[a] $p = .05$. *$p < .05$. **$p < .01$.

Although reading comprehension at T2 accounted for a large proportion of variance in both groups (bilingual final β = .59, $p < .01$; monolingual final β = .31, $p = .03$), two additional predictors had significant regression weights in the final equations for each group. In the bilingual group, both German vocabulary at T1 and word decoding at T2 were significant predictors of the growth in reading comprehension. For the monolingual group, however, decoding and phonological awareness at T2 were the two significant predictors of growth in reading comprehension[10]. In the monolingual group, German vocabulary contributed no variance to the final regression equation, whereas phonological awareness made no significant contribution in the bilingual group. This pattern fit well with Hypotheses 3c and 3d, which anticipated that growth in reading comprehension would be differentially predicted for the two groups, with German vocabulary playing a stronger role for the bilingual group and phonological awareness playing the stronger role for the monolingual group. To test whether these differences were statistically significant between the two groups and detect significant differential patterns in the predictive powers of German vocabulary and phonological awareness, interaction terms were integrated into the equation.

The two variables that clearly differed between the monolingual and bilingual groups in the separate regression equations (phonological awareness at T2 and German vocabulary at T1) were selected to test for significant differences with interaction terms. The results of the hierarchical multiple regression with the two interaction terms are shown in Table 45. With reading comprehension at T3 as the criterion variable, and reading comprehension at T2 as the first step in the regression equation, the variables that contributed additional predictive power to the growth of reading comprehension skills from second (T2) to third (T3) grade could be estimated. The linear combination of the predictor variables with all eight steps was highly significantly related to reading comprehension at T3, $R^2 = .67$, $F\ (10, 124) = 28.31$, $p < .01$. Five of the nine predictor variables (other than reading comprehension at T2) significantly accounted for the remaining variance in the growth equation: German vocabulary at T1, decoding at T2, phonological awareness at T2, group membership, and the interaction between phonological awareness at T2 and group[11].

[10] Separate hierarchical multiple regressions for each group conducted with reading comprehension gain scores as the criterion variable to represent growth produced similar results with regard to the unique amount of explained variance for each predictor; however the overall explained variance for the gain score was, as would be expected, much lower without accounting for performance on the reading comprehension measure at T2 (TB final adjusted R^2= .07, GM final adjusted R^2 = .15).

[11] A regression conducted with gain scores also produced similar results in the combined regression equation with interaction terms. The overall explained variance was, however, much lower without accounting for performance on the reading comprehension measure at T2 (adjusted R^2 = .10). For a discussion of the difficulties of using gain scores in similar analyses see Hendrickson and Jones (1987).

Table 45

Stepwise Hierarchical Multiple Regression Showing the Contribution of German Vocabulary, Phonological Awareness, and Word Decoding with Interaction Terms at all Times of Measurement to Growth in Reading Comprehension Performance for all Participants

Step		Adjusted R^2	ΔR^2	Final β
1	Reading comprehension T2	.57	.57**	.48**
2	Decoding T1	.57	.01	-.09
3	German vocabulary T1	.60	.03**	.24*
4	PA T1	.62	.03**	.10
5	Decoding T2	.65	.03**	.24**
6	German vocabulary T2	.65	.00	-.04
7	PA T2	.65	.01	.31**
8		.67	.03*	
	Language group (1= Bilingual, 0 = Monolingual)			.53**
	Interaction term: PA T2 x language group			-.56**
	Interaction term: Vocabulary T1 x language group			-.04

Note. PA = Phonological awareness. Dependent variable was reading comprehension at T3. Significance levels for the change in R^2 refer to the significance in the change in *F* values for each entered step of the regression. For purposes of simplification, the *F* values are not displayed here, only R^2 values. Both the interaction term and the relevant variable (here PA and German vocabulary) were *z*-transformed to avoid the interaction terms confounding with the original variables (PA or German vocabulary).
$^*p < .05$. $^{**}p < .01$.

This analysis provides strong evidence that Hypothesis 3c may be an accurate depiction of the processes underlying reading skills among bilingual and monolingual beginning readers. The strong negative regression coefficient for the phonological awareness interaction term indicates that phonological awareness was indeed a significantly weaker predictor of growth in reading comprehension abilities for the Turkish bilingual group than it was for the German monolingual group. Conversely, the insignificant German vocabulary interaction regression coefficient provided no support for Hypothesis 3d, which predicted German vocabulary would play a stronger role in reading development for the bilingual group.

4.5. Model fit

This section utilizes structural equation models (SEM) to explore the component abilities thought to be responsible for developing reading comprehension skills in the bilingual and monolingual groups of children. The a priori models tested are based on the modified model of

German reading proposed by Näslund and Schneider (1991), as well as on the evidence provided by the multiple regression analyses from this investigation. Because some authors (e.g., Jaccard & Wan, 1996) recommend that studies with smaller sample sizes utilize multiple regression methods (as performed in Sections 4.3. and 4.4) as opposed to structural equation modeling, these analyses are presented here under the premise of tentative first analyses and must be replicated with substantially larger samples in the future. Nonetheless, the models provide additional valuable insights and should not be disregarded altogether due to small sample sizes.

The use of models that integrate latent constructs have the advantage over multiple regression analyses of accounting for measurement error; an attribute of particular benefit when not all instruments perform with high levels of reliability. For that reason, latent variables were used whenever possible in the following models. The four subscales of the phonological awareness tasks were used as manifest variables, serving as indicators for the latent construct of *phonological awareness abilities* at each time of measurement. Similarly, the subscales of the German vocabulary tests (two at T1, three at T2) made up the latent construct representing *German expressive vocabulary*. Because over one-hundred items measured decoding and because it demonstrated substantial reliability as a measurement instrument, word decoding was incorporated into the models as a manifest variable at both times of measurement. The same rationale was used in allowing the twenty-item reading comprehension measures to be entered into the model in manifest form for both T2 and T3.

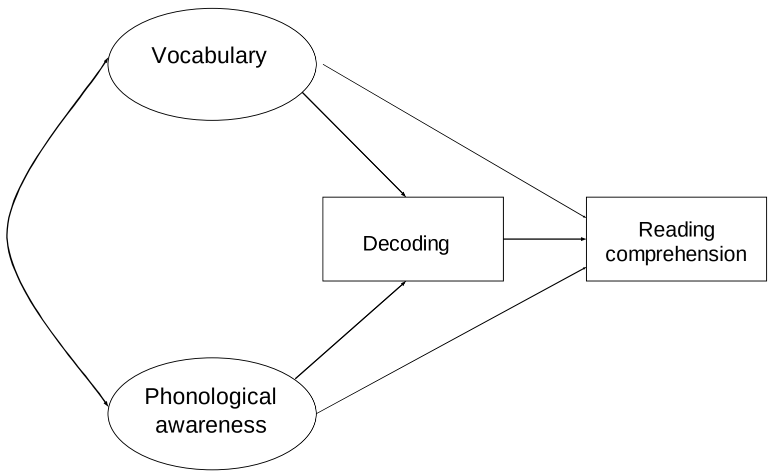

Figure 10. Structural model based on Näslund & Schneider (1991) modified for analyzing predictors of reading comprehension with structural equation modeling

Since removing extraneous variables also aids the parsimony of the models, extraneous variable were removed whenever possible. Based on the correlation and regression analyses in the previous two sections, the decision was made to eliminate the verbal memory variables from the Näslund and Schneider (1991) model. Cognitive abilities and listening comprehension were also excluded from the models. No analysis in this investigation thus far has confirmed the importance of any of those three abilities in predicting reading skills directly or indirectly in the present sample. Figure 10 illustrates the final revised version of the model to be tested. Ovals represent latent variables, whereas rectangles represent the manifest variables.

In that the following models were intended to explain reading development, time was dealt with carefully in the explanation of causal effects. More specifically, three guidelines set by Gollob and Reichardt (1987) were adhered to. First, since causation happens over time, the value of a criterion variable can only be caused by the values of a prior variable. Second, prior measures of the dependent variable are included in order to control for the possibility of autoregressive effects. Autoregressive effects are examined specifically in the longitudinal growth models at the end of this section (Figure 14). Finally, as the size of an effect is often dependent on the length of time between the hypothesized cause and effect, it was considered important that the measures were set at equal intervals.

The analyses proceeded in a series of steps based loosely on the procedures conducted in a similar study by Verhoeven (2000) and the recommendations of Garson (2003) and Loehlin (1998). First, the modified Näslund and Schneider (1991) model is fit to both groups together with the aim of determining a common solution. Since no common solution was possible for the two groups together, the model was tested separately for the data from the German monolingual and Turkish bilingual children (see Verhoeven, 2000)[12]. The model was estimated with a maximum likelihood method and modified for each group by removing negative and insignificant paths (only when clearly interpretable) until a viable solution was found[13]. Ideally, the model would have been cross-validated with a second sample, but that final confirmatory step was not possible due to the restricted sample size available.

The analyses were conducted with two distinct sets of models. In the first set, simple predictive models were tested for each measurement combination as was done with the regression equations above (T1 variables predicting T2 reading comprehension, T1 variables predicting T3 reading comprehension, T2 variables predicting T3 reading comprehension) to test Hypotheses 4a and 4b. In essence, the first set of models investigates the respective roles of phonological awareness and verbal abilities regarding their role in the development of reading comprehension skills at a later point in time. In order to test Hypotheses 4c and 4d, however, a

[12] In one case, it was possible to find a common model for both groups. Predicting reading at T3 was successful with the same combination of variables from T2 for the Turkish bilingual and German monolingual data. This will be discussed in detail below.

[13] Non-significant paths were removed from SEM analyses both for ease of interpretability and based on findings showing that non-significant paths often inflate standard errors (see Cohen, Cohen, West, & Aiken, 2003).

second set of models was developed in which growth is examined under consideration of possible autoregressive effects. The second, more intricate set of models aims at analyzing the combination of variables at both T1 and T2 responsible for growth in reading comprehension skills by controlling for reading comprehension at T2 as well.

Six goodness-of-fit indices (chi-square, CMIN/DF, CFI, NFI, IFI, and AIC) were selected to be reported for the following models. Due to the large and diverse discussion currently taking place with regard to fit indices and appropriate cutoffs (e.g., see Kelloway, 1998; Tanaka, 1993), a range of different fit indices were chosen. An explanation of the selected fit measures with recommended cutoff values is provided in Appendix D as a reference for the reader.

H4a & H4b: As a continuation of H2c and H2d, it is expected that phonological awareness will play a weaker role among the bilingual group, whereas verbal abilities will play a greater role among the bilingual group than the monolingual group at each time of measurement with the proposed structural equation model (H4a); the proposed model will be a substantially poorer fit for the bilingual group than for the monolingual group. A larger amount of variance will be left unexplained for the bilingual group (H4b)

Following the modeling process described above, data from the Turkish bilingual and German monolingual groups were tested separately to determine if phonological awareness played a stronger role in predicting reading comprehension for the German monolingual group, while German vocabulary played a stronger role in predicting reading for the Turkish bilingual group. The structural equation models calculated in this section mirror the multiple regression models described in Sections 4.3. and 4.4., but with three primary additional advantages: 1) phonological awareness and verbal abilities can be examined as latent variables, thus accounting for measurement error, 2) the Näslund and Schneider (1991) model can be tested as proposed with mediator effects, and 3) autoregressive effects can be controlled for. Figures 11, 12, and 13 schematically represent the theoretical model with the corresponding regression weights depicted for each significant path. Non-significant or irrelevant paths are marked with “ns”. Table 46 provides an overview of the fit of all longitudinal models predicting reading comprehension performance at the end of second grade (T2) and the middle of third grade (T3) with the six fit indices as well as the total amount of variance accounted for in the criterion variable.

Table 46

Fit Indices for Three Different Longitudinal Combinations of Structural Equation Models for the German Monolingual and Turkish Bilingual groups

Model	χ^2	*df*	*p*	CMIN/DF	CFI	NFI	IFI	AIC	Final R^2
T1 to T2									
Monolingual	14.35	8	.07	1.79	.95	.96	.91	52.35	.49
Bilingual	30.93	14	.03	1.72	.94	.95	.89	82.93	.47
T1 to T3									
Monolingual	14.47	8	.07	1.81	.95	.95	.90	52.47	.61
Bilingual	15.48	18	.22	1.29	.98	.98	.92	61.84	.50
T2 to T3									
Monolingual Bilingual	22.73	16	.12	1.42	.97	.92	.97	98.73	.81 .44

Note. In the final model (T2 to T3), a common fit was found for both groups. Fit indices are therefore provided for the common model, but the R^2 is still presented for both groups separately.

The first set of models represents the latent constructs of vocabulary and phonological awareness and the manifest variable decoding in the middle of second grade with their relationships to one another in predicting reading comprehension at the end of second grade. Because different paths were meaningful for the two groups (for example, the path between phonological awareness and reading comprehension was non-significant for the Turkish-German bilingual group, but was highly significant for the German monolingual group), the computation of two separate models was required. Figure 11 gives a schematic representation of the SEM for the both groups.

For the bilingual group, three significant regression paths and one covariance path are depicted. The effect of phonological awareness on reading comprehension was partially mediated through decoding, that in turn, significantly predicted reading. Additionally, German vocabulary contributed significantly to predicting reading comprehension at T2. With data from the Turkish bilingual group, this model accounted for 47% of the variance in reading comprehension at T2. The model fit for the Turkish bilingual data set was reasonable, with a CMIN/DF value of under 2 and a CFI of .94. The chi-squared value was, however, significant (p = .03).

With the data from the German monolingual group, only three significant paths emerged. Phonological awareness had both a direct significant effect on reading comprehension and an effect mediated over decoding. Vocabulary had no direct or indirect effect on reading comprehension in the analyses with the monolingual group data. For both the Turkish bilingual and German monolingual groups, the overall R^2 was similar with explained variances of 47% and 49% respectively. The model fit with the German monolingual data was good with a CMIN/DF under 2, a CFI of .95, and a non-significant chi-square (p = .07).

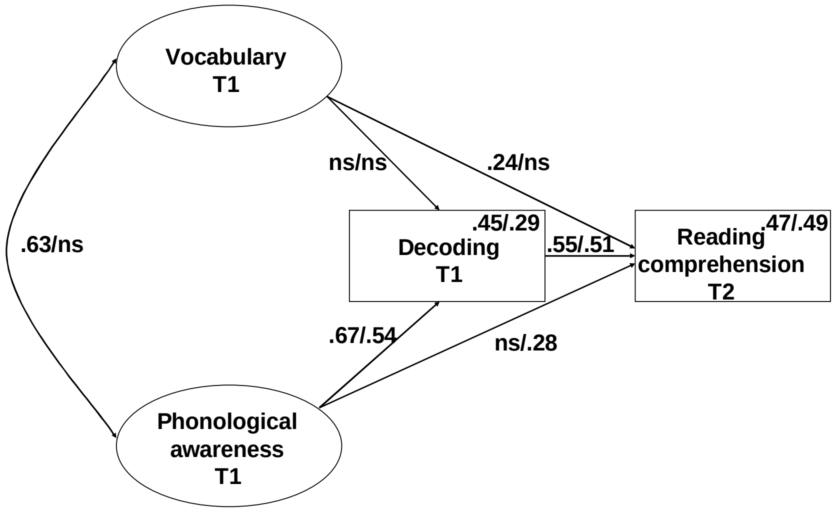

Figure 11. Structural equation model with mid-second grade predictors (T1) for reading comprehension at the end of second grade (T2) for the Turkish bilingual group/German monolingual group

Note. Coefficients printed in the variable boxes represent the R^2 for that variable.

The analyses of the Turkish bilingual and German monolingual data for T1 predictors of T2 reading comprehension provide support for Hypotheses 4a and 4b. As predicted by Hypothesis 4a, significant regression weights were found from vocabulary to reading comprehension for the bilingual group, but not for the monolingual group. Conversely, as anticipated, T1 phonological awareness had a direct effect on T2 reading comprehension for the German monolingual group, but not for the Turkish bilingual group. With regard to model fit, as predicted by Hypothesis 4b, the model appears to fit better the monolingual data set better than the bilingual data set for the T1 predictors of T2 reading comprehension. This is indicated by the much smaller chi-square value for the monolingual model (14.35 vs. 30.93) as well as the substantially lower AIC value (52.35 vs. 82.93).

Comparing the SEM analyses to the multiple regression analyses in Section 4.3. (Table 39) exposes several small discrepancies. First, in the multiple regression analysis, the overall R^2 for the German monolingual group was larger than the Turkish bilingual group (.58 vs. .47); the amount of explained variance in the SEM analyses was essentially equal for both groups. Second, the multiple regression analyses showed that phonological awareness at T1 was an important predictor of reading comprehension at T2 for both groups. In the SEM analyses, a direct path was only found in the German monolingual group, whereas both groups showed phonological awareness at T1 to be mediated over decoding at T1 to predict reading comprehension at T2.

Finally, it is interesting to note that the multiple regression analyses did not find vocabulary to contribute significantly to reading comprehension in either group. In the SEM analyses, the latent variable of vocabulary at T1 significantly contributed to reading comprehension at T2. Possible reasons for these discrepancies will be discussed in the Discussion section below.

Figure 12 provides the regression weights for the models examining the relationships between the latent and manifest variables in the middle of second grade (T1) and reading one year later in the middle of third grade (T3). With a few incongruities, these analyses produced a similar pattern of regression weights and explained variance as found in the previous analyses for predicting reading comprehension at T2.

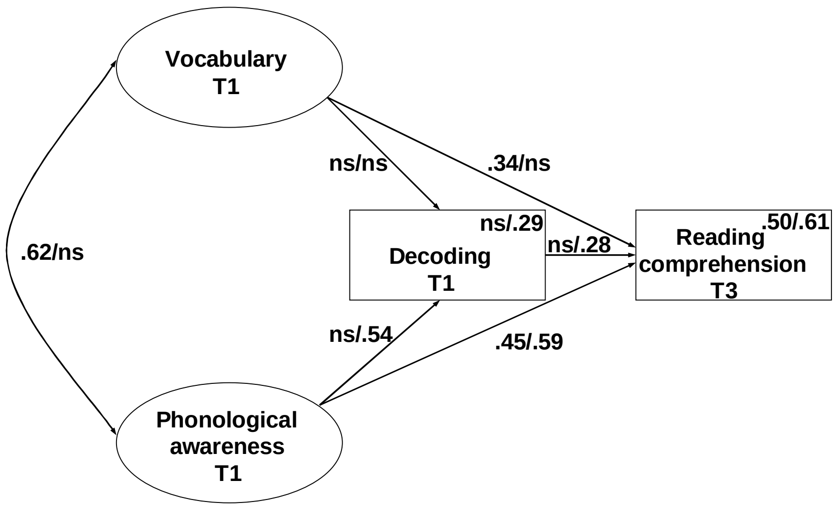

Figure 12. Structural equation model with mid-second grade predictors (T1) for reading comprehension at the end of second grade (T3) for the Turkish bilingual group/German monolingual group

Note. Coefficients printed in the variable boxes represent the R^2 for that variable.

For the Turkish bilingual data, both latent constructs, vocabulary and phonological awareness at T1, had a direct effect on reading comprehension at T3. T1 Decoding no longer plays a mediating role between T1 phonological awareness and T3 reading comprehension as it did in the prediction of reading comprehension at T2. Instead, T1 phonological awareness is a significant predictor of T3 reading comprehension. However, the T1 predictors explained a similar amount of variance in T3 reading comprehension as T2 reading comprehension (50% and 47% respectively). In contrast to the model of T1 to T2 variables in the Turkish bilingual data set, when investigating Turkish bilingual participants' reading comprehension at T3, decoding no longer played a significant predictive role. Unlike the T1 to T2 analysis, this model resulted in a low CMIN/DF ratio (1.29), a very high CFI value (.98), and a strongly non-significant chi-square fit for the bilingual group (p = .22) indicating an excellent model fit.

The data from the German group for the T1 to T3 model also produced similar results to the T1 to T2 analysis. Again, the latent vocabulary construct demonstrated no significant regression weights onto decoding or reading comprehension. It appears that phonological awareness and decoding abilities carried the largest predictive power for the monolingual group for this model as well. Somewhat different to the T1 to T2 analyses, the amount of explained variance for T3 reading comprehension was moderately higher (61% compared to 49% for T2 reading comprehension).

In general, the regression weights in the SEMs predicting T3 reading comprehension from latent verbal constructs and word decoding at T1 supported Hypotheses 4a and 4b. Hypothesis 4a was supported by the fact that the T3 reading comprehension predicted with T1 data from the Turkish bilingual group was more dependent on the predictive power of vocabulary, whereas T1 phonological awareness carried somewhat more weight in predicting reading comprehension in the model computed with the monolingual data. This indicated that, as predicted by Hypothesis 4a, phonological awareness played a larger role in reading comprehension for the monolingual participants than for the bilingual participants[14], and conversely, that vocabulary is a significant predictor of reading comprehension for the bilingual group, but not for the monolingual group. The T1 to T3 SEM analyses also provided support for Hypothesis 4b, indicating that the model fit better for the German monolingual group. The data from the German monolingual group both explained a larger amount of variance (R^2 = .61) compared to the Turkish bilingual group (R^2 = .50) and resulted in a lower (better) AIC fit score for the monolingual group (52.47 vs. 61.84 for the bilingual group).

Compared to the parallel multiple regression analyses presented in Section 4.3. (Table 39), the T1 to T3 SEM analyses differed in two regards. First, the multiple regressions showed T1 decoding to be a significant predictor of T3 reading comprehension for the bilingual group. This was not substantiated by the SEM analyses with the bilingual group data, for which decoding showed no significant power to predict T3 reading comprehension. Secondly, the multiple regression analyses indicated no significant role for T1 vocabulary in predicting T3 reading comprehension for either group. The SEM analyses, however, found the latent variable of T1 vocabulary to be a significant predictor of T3 reading comprehension for the bilingual group.

The final longitudinal SEM analysis, in which the latent verbal constructs and decoding from late second grade were used to predict reading in mid-third grade, produced comparatively

[14] A statistical difference between the regression weights for phonological awareness (bilingual = .45, monolingual = .59), however, cannot be tested for significant differences since the models were calculated separately.

different effects. For this analysis, a single model was found to fit both groups; the same paths demonstrated significance in both groups. Although the fit indices for this model are somewhat weaker than as in the other models, the CFI, NFI, IFI, and CMIN/DF values appear to indicate an acceptable model (see Table 46). Neither the bilingual nor the monolingual data demonstrated a significant path from vocabulary at T2 to reading comprehension at T3. For both groups, strong relationships were revealed between phonological awareness and word decoding.

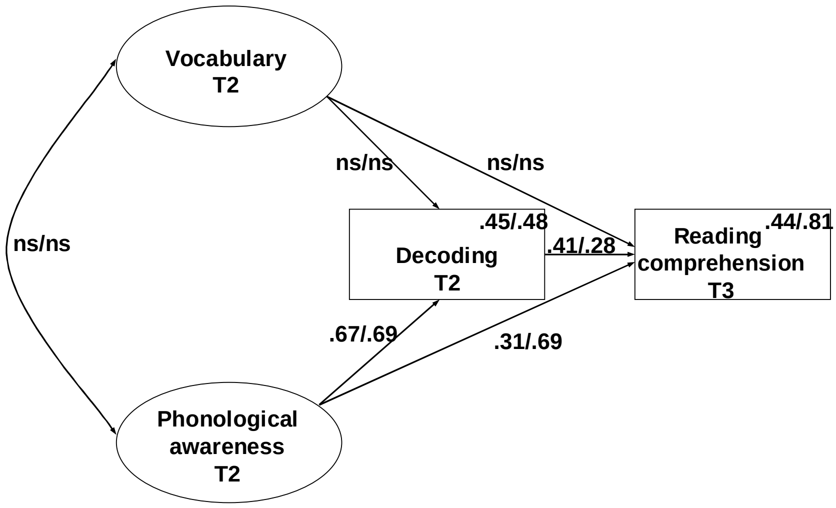

Figure 13. Structural equation model with mid-second grade predictors (T2) for reading comprehension at the end of second grade (T3) for the Turkish bilingual group/German monolingual group

Note. Coefficients printed in the variable boxes represent the R^2 for that variable.

Although direct paths led from the phonological latent construct to reading comprehension for both groups, the regression weight for the bilingual group (r = .31), appeared substantially lower than that of the monolingual group (r = .69). The path was thus tested for significant differences. By holding the regression weight constant for both groups and retesting the model, it was found that forcing the path from latent phonological awareness at T2 to reading comprehension at T3 to be equal for both groups resulted in a significantly poorer fit, CMIN(1) = 4.37, p = .04. It can therefore be inferred that the latent construct for phonological awareness at T2 is a significantly stronger predictor of reading comprehension at T3 for the German monolingual group than for the Turkish bilingual group. This model provides strong support for the phonological aspect of Hypothesis 4a (the prediction that phonological awareness would be a weaker predictor of reading comprehension for the bilingual group), but no evidence of the vocabulary aspect of the hypothesis (that vocabulary would be a stronger predictor of reading comprehension for the bilingual group). Hypothesis 4b, that predicted a better fit for the monolingual group, was clearly supported in the large discrepancy between explained variance in the German monolingual group ($R^2 = .81$) compared with ($R^2 = .44$). The fact that the AIC fit index was substantially poorer for the combined bilingual/monolingual model for the T2 to T3 predictors (AIC = 98.73) compared to the separate models calculated for the other measurement combinations (AIC range 52.35-82.93), could also be interpreted as evidence that the combined model was a poorer fit than the separate models-- and poorer than the individual monolingual models in particular.

Compared to the multiple regression models examining the T2 predictors of T3 reading comprehension, the SEM analyses were reasonably similar. In both the multiple regression analyses and the T2 to T3 SEM analyses, T2 decoding and T2 phonological awareness appeared to be the strongest predictors of T3 reading comprehension for the monolingual group. Although the multiple regression analyses showed T2 decoding to be the only significant predictor of T3 reading comprehension for the bilingual group, when the data was examined with latent and mediating variables in the SEM analyses, both T2 decoding and T2 phonological awareness emerged as significant predictors of T3 reading. A very clear similarity between the multiple regression analyses and the SEM analyses was apparent in the sizeable difference in total R^2 that was found between the two groups. In both analyses, the amount of variance explained by the model for the Turkish bilingual group was only about half of the variance explained for the monolingual group (multiple regression: TB = 38%, GM = 70%; SEM: TB = 44%, GM = 81).

Overall, the three sets of structural equation models provided relatively consistent evidence for Hypothesis 4a. Stronger predictive power was found for the latent phonological awareness variable with data from the monolingual group, whereas the latent German vocabulary abilities variable was more frequently found to be a significant predictor of reading comprehension in the monolingual group. The AIC fit indices also persistently signified better model fits for the monolingual group, thus supporting Hypothesis 4b.

H4c & H4d: As a continuation of H3c and H3d, phonological awareness will play a lesser role among the Turkish-German bilingual group, whereas verbal abilities will play a greater role among the bilingual group than the German monolingual group in the explanation of growth in reading comprehension with proposed structural equation model (H4c); the proposed model of reading for growth in reading comprehension will not fit a Turkish-German bilingual population as well as a monolingual German population. A larger amount of variance will be left unexplained for the bilingual group (H4d).

The final set of SEM analyses aimed to test a model of reading development based on the Näslund and Schneider (1991) model of reading, modified according to the empirical results already presented from this investigation and adapted to account for growth. Similar to the previous sections, it was hypothesized that phonological awareness would be a stronger predictor of reading growth for the monolingual group, while German vocabulary would have a greater impact on reading development among the bilingual group. These analyses are a continuation of the multiple regression models in Section 4.4. (Table 44) aimed at predicting growth in reading comprehension abilities from second (T2) to third grade (T3). Compared to the multiple regression analyses in Section 4.4., these analyses have the advantage of taking measurement error in the phonological awareness and vocabulary measures into account by investigating their impact on the dependent variable as latent variables. Additionally, these SEM analyses are better tools for controlling for autoregressive and mediating effects. Again, the extraneous variables were removed to increase parsimony. Figure 14 displays a representation of the base model.

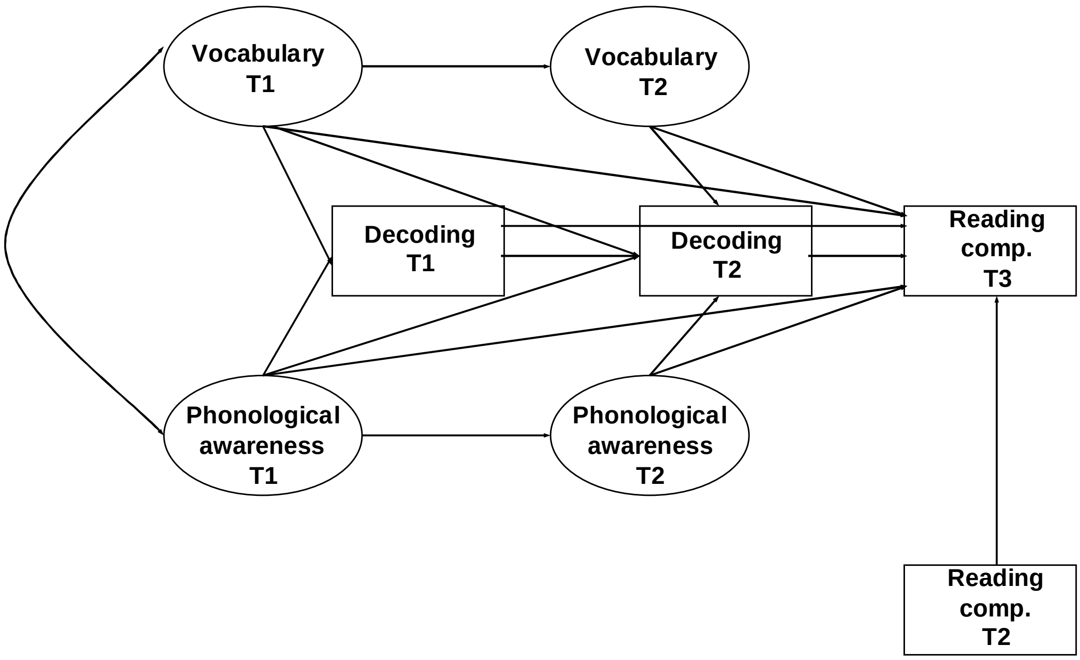

Figure 14. Structural model based on Näslund & Schneider (1991) modified for analyzing predictors of growth in reading comprehension with structural equation modeling

Since no common fit could be found based on the a priori model, separate structural equation models were computed for each group with the latent and manifest variables from both measurement times in second grade (T1 and T2) and with reading comprehension in third grade (T3) as the final dependent variable. To test for growth, the manifest variable for reading comprehension at T2 was separately regressed onto the T3 variable for reading, leaving only the six-month growth variance unaccounted for. Table 46 shows the goodness-of-fit indices for each group along with the total amount of variance in reading comprehension at T3 accounted for by the complete model.

Table 47

Fit Indices for the Longitudinal Growth Models for the German Monolingual and Turkish Bilingual Groups

	χ^2	*df*	*p*	CMIN/DF	CFI	NFI	IFI	AIC	Final R^2
Monolingual	141.49	51	<.01	2.77	.76	.69	.78	219.49	.65
Bilingual	153.94	73	<.01	2.11	.85	.76	.86	245.94	.59

Figure 15 presents the modified growth model for both groups together. For both groups, only seven of the fourteen paths were found to be significant predictors of reading comprehension growth. Three of the fourteen paths differed, however, for each group.

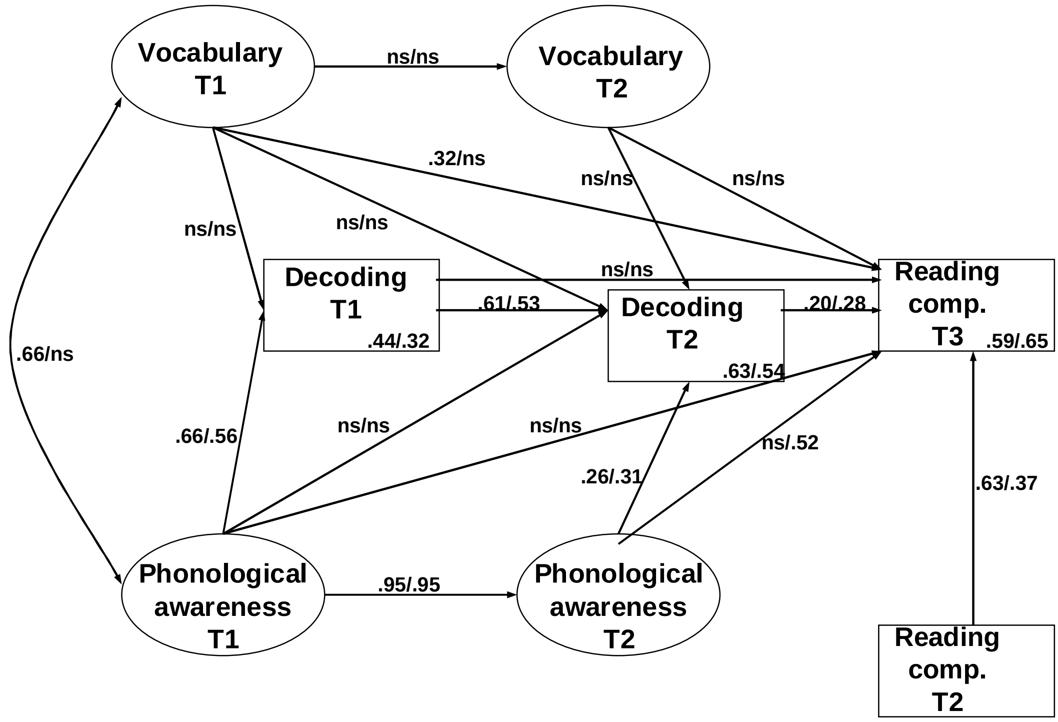

Figure 15. Structural equation model with all variables for predicting growth in reading comprehension for the Turkish bilingual group/German monolingual group

Note. Coefficients printed in the variable boxes represent the R^2 for that variable.

The Turkish bilingual group's model will be discussed first. After controlling for reading comprehension at T2, two variables demonstrated direct effects on reading comprehension at T3: the latent German vocabulary construct at T1 and word decoding at T2. Phonological awareness at both T1 and T2 produced indirect effects, mediated by word decoding at each point of measurement. The results of this reading comprehension growth analysis are very much in line both with the multiple regression analyses in Section 4.4. in which reading comprehension at T2 is accounted for in the prediction of reading comprehension at T3 (see Table 44), as well as the separate SEM analyses presented at the beginning of this section (Figures 11 - 13). However, with the limited sample size and larger number of variables, the less parsimonious model resulted in a poorer fit with a significant chi-square (χ^2 (73)= 153.94, $p < .01$) and weaker scores on the remaining fit indices (CMIN/DF = 2.11, CFI = .85).

For the German monolingual group, only six of the 14 potential paths demonstrated significance. As predicted, phonological awareness as a latent construct was strongly related to the development in reading with both direct (T2) and indirect paths (mediated by decoding at both T1 and T2). In accordance with the H4c prediction, the latent construct for vocabulary produced no significant regression weights onto reading or decoding at any of the measurement points. These SEM analyses of growth in reading comprehension from T2 to T3 are generally consistent both with the multiple regression analyses in Section 4.4. (see Table 44) and the individual SEM analyses at the beginning of this section (Figures 11-13). According to the chi-square value (χ^2 (51)= 141.49, $p < .01$), the model was not statistically adequate, and the remaining indices also gave a weak impression of the model's fit (CMIN/DF = 2.77, CFI = .76). However, with four latent variables and a sample of this size, this is not surprising.

Overall, the structural equation models for growth provided support for Hypothesis 4c, in that phonological awareness clearly played a more important role in predicting reading comprehension when analyzed with the German monolingual data set. Furthermore, only within the bilingual data was a direct path found from (T1) German vocabulary onto reading comprehension. Evidence was also found for Hypothesis 4d. As expected, the model was able to predict more variance for the monolingual group than for the bilingual group. Also as expected, the AIC index showed poorer model fit for the bilingual group. In light of the weak goodness-of-fit indices and small sample size, these findings should be interpreted only as exploratory indications of possible patterns of reading comprehension growth prediction for bilingual and monolingual children.