The Language Proficiency of Turkish Returnees from Germany: An Empirical Investigation of Academic and Everyday Language Proficiency

Helmut Daller
Faculty of Languages and European Studies, University of the West of England, Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY, UK

Rüdiger Grotjahn
Ruhr-Universität Bochum, Seminar für Sprachlehrforschung, 44780 Bochum, Germany

This study analyses the everyday and academic language proficiency of German–Turkish bilinguals who grew up in Germany as children of Turkish immigrants and learned German as a second language. They attended German schools before returning to Turkey, where German formed part of the school curriculum. The bilinguals of our sample were all university students of German philology. Although the returnees’ command of German seemed to be native-like at a surface level, their teachers and lecturers frequently complained about their academic language proficiency. Therefore, their everyday and their academic language proficiency was measured with two different C-tests, developed on the basis of Cummins’ distinction between basic interpersonal communicative skills (BICS) and cognitive-academic language proficiency (CALP). The scores obtained were compared with those of a control group of Turkish university students who had spent their childhood in Turkey and learned German at school. Statistical analysis of the data yielded the following results: (1) the C-tests were a useful means of measuring different aspects of language proficiency; (2) the returnees showed significantly higher scores for everyday language proficiency in German compared to the control group; (3) the main predictor variables were school type and length of time spent at a German school; and (4) attrition of everyday language proficiency in German depended on the length of time since return to Turkey.

Returnees and Migration

The Turkish migration to Germany started in the early 1960s. From the beginning of this migration process there was also a constant ‘re-migration’ of Turks back to their home country. Among these returnees there are thousands of Turkish children who return every year during their school career to their home country (Zentrum für Türkeistudien, 1994: 25).

The informants in our sample were either born in Germany or had gone there at a very young age. The majority of these children acquired Turkish as a first language before learning German at school. The first language was not supported at the German school in most cases and therefore German became the only language for academic purposes. In our sample the average age at return is 13 years which is representative for returnees in general (see Daller, 1995, 1999). After returning to Turkey the informants continued their education in Turkish,
which meant a second break in their language development. Turkish was the main means of everyday communication in Germany but subsequently became the most important school language in Turkey.

Special schools with a bilingual programme were established to support the (re)integration of these children (the so-called Anadolu schools). Many of the returnees stayed in contact with the German language after their return to Turkey, either at these Anadolu schools or at the university.\(^1\) Many Turkish teachers at the Anadolu schools complain about the lack of academic language proficiency of the returnees in both languages, whereas their everyday language proficiency seems to be native-like. A wide range of research data (for an overview see Reich, 1995, as well as Treffers-Daller & Daller, 1995) confirms this assessment by the teachers. In this article we focus on everyday and academic language proficiency\(^2\) in German (for a comparison between the German and Turkish proficiency see Daller, 1995, 1999).

**Theoretical Background**

The present study derives from the theories of Jim Cummins. In different studies Cummins analysed the language proficiency of migrant children in North America. For these immigrants Cummins showed the existence of a discrepancy between seemingly native-like language proficiency at a surface level and poor academic language proficiency in comparison with their monolingual peers. He developed a framework, which distinguishes between *basic interpersonal skills* (BICS) and *cognitive–academic language proficiency* (CALP). Later he referred to this dichotomy also as *conversational and academic language proficiency* (see Cummins, 1984, 1991). Migrant children reach the proficiency level of their monolingual peers in BICS quite quickly but need much more time to get to their level in CALP.

Age appropriate BICS can be achieved within two years after immigration whereas for CALP five, seven and even ten years are mentioned in the literature (Cummins, 1981: 148; Skutnabb-Kangas, 1984: 113; Collier, 1989: 529). This difference in the time necessary to achieve appropriate language proficiency in BICS and CALP leads to misinterpretations of a child’s real linguistic abilities. Cummins (1984: 4) speaks of a ‘linguistic façade’ in many migrant children. The seemingly native-like fluency in face-to-face oral communication leads to the false conclusion that the child has age-appropriate language proficiency at academic level as well. Failure at school is then erroneously attributed to the child’s general abilities and not to delayed language acquisition at academic level.

In answer to critics Cummins and Swain (1983: 36) admit that the dichotomy BICS versus CALP is too simplistic. Cummins (1984) therefore proposed a new ‘framework of language proficiency’ which is two-dimensional. The model is displayed in Figure 1.

This model describes language proficiency with the help of the two dimensions *cognitively undemanding* versus *cognitively demanding* and *context-embedded* versus *context-reduced*. The model can also be used to describe the nature of specific language tasks. Quadrant A of the new model represents a type of language proficiency which is closely linked to the old label BICS whereas CALP can be identified with quadrant D (see also Hacquebord, 1989: 51; Verhallen, 1994: 63).
The Study

Sample

The experimental group consisted of 159 Turkish returnees. All were university students of German philology at universities in Turkey at the time of the investigation (1993–1994). Their average age was about 20 and they had been back in Turkey for about seven years, which means that their average age at return was about 13. The informants had either been born in Germany or had gone there when very young. We excluded all persons who had come to Germany at the age of three or more in order to keep the sample as homogeneous as possible and to reduce the number of variables. The control group consists of 10 Turkish university students of German philology at the same university. They had learned German as a foreign language in Turkey. They had only had contact with the target language inside the classroom, apart from a few short stays in Germany.

Tests

We used two different C-tests\(^3\) and a questionnaire with items on social background and language acquisition biography. The C-tests were developed on the basis of two sets of ten texts. One set was taken from course books and other publications in German philology, the second one from newspapers about everyday topics such as traffic jams and holidays. A group consisting of seven experienced teachers of German as a foreign language rated these 20 texts as written in either ‘academic’ or ‘everyday language’. We chose the four texts classified as most clearly academic (all from the German philology literature) and four texts rated as most clearly everyday language. On the basis of these two sets of four texts, we then constructed an academic language proficiency C-test (C-test ALP) and an everyday-language proficiency C-test (C-test ELP).

University lecturers administered these C-tests to the students as paper and pencil tests in group sessions during the teaching hours. The texts of the tests are given in the appendix. The underlined parts were blanks in the original, which the students had to fill in. The given time limit was 90 minutes (one teaching bloc).
which did not cause too much time pressure for the students. We applied exact scoring, which means that there was no weighting between different types of error.

C-texts and CALP

Some authors argue that the C-test as a written test cannot cover the BICS aspect of language proficiency because BICS are closely related to oral language use (see, e.g. Grotjahn & Tönshoff, 1992). In order to avoid misinterpretations we therefore decided not to use the terms BICS and CALP but everyday language proficiency (ELP) and academic language proficiency (ALP). One of our C-tests is designed to measure ELP, the other to measure ALP. Our distinction can still be interpreted within the framework of Jim Cummins as shown in Figure 2. It is therefore an attempt towards an operationalisation of this framework. Figure 2 differs from Cummins’ original version (Figure 1) in the way that quadrant D which represents CALP is larger than the other three quadrants. This reflects the fact that the acquisition of CALP is a more demanding task than the acquisition of BICS and takes therefore much more time (see Theoretical Background, above).

Both C-tests can be placed into quadrant D of Figure 2, which means that they both measure some academic proficiency. Nevertheless C-test ALP is far more oriented towards academic language proficiency and C-test ELP comes close to quadrant A, which represents BICS.

Most authors claim that the C-test in its traditional form (that is with a neutral text basis) measures mainly one single factor: general language ability (Klein-Braley, 1994; Grotjahn, 1995). In addition to this we tried to measure academic and everyday language proficiency in the present study. This means that a proficiency model based on the C-tests of the present study includes: General Language Proficiency (GLP), Academic Language Proficiency (ALP) and Everyday Language Proficiency (ELP). Our test results also include a Standard Error of Measurement (SE) and a Guessing Factor (GF), since some of the blanks filled in can be due to mere guessing. It is thus assumed that the observed scores in the two C-tests are the result of the following factor constellations:
Scores C-test ALP = GLP + ALP + SE + GF
Scores C-test ELP = GLP + ELP + SE + GF

It has to be investigated whether the factor GLP (General Language Proficiency) is partly identical with ELP (Everyday Language Proficiency).

Hypotheses

The aim of the present study is to investigate the following hypotheses:

- It is possible to develop C-tests that distinguish between everyday and academic language proficiency.
- Returnees show a high level of everyday language proficiency in comparison with the control group. There is, however, no significant advantage of the returnees regarding academic language proficiency.
- The language proficiency of the returnees can partly be predicted on the basis of their language acquisition biography.

Results

C-test data

We will now statistically assess whether our C-test data confirm the hypothesis of two different dimensions of L2 ability (see above). If two-dimensionality is not confirmed, then there are no grounds for distinguishing between academic and everyday language proficiency on the basis of the test scores. Subsequently, the reliability of the C-tests used will be examined. In the light of the results of the dimensionality and reliability check it will then be investigated whether the number of C-test texts necessary for the measurement of L2 ability might be reduced. Since the focus of the present article is on the language ability of Turkish returnees, we will restrict ourselves to this group (n = 159) in all subsequent analyses.

Assessment of dimensionality

There exist many methods for the examination of the dimensionality of a measurement instrument. We used two different approaches, namely factor analysis and the classical latent additive test model. All calculations were carried out with the help of SPSS (PC version).

Factor analysis

The C-test for everyday language proficiency (C-test ELP) consists of four parts based on four everyday texts, and the C-test for academic language proficiency (C-test ALP) consists of four parts based on four academic texts. In the following computations these parts are called elp1, elp2, elp3 and elp4 for the C-test ELP and alp1, alp2, alp3 and alp4 for the C-test ALP. The matrix of intercorrelations among these eight C-test texts is shown in Table 1. It can be seen from Table 1 that overall the correlation both within the ELP-group and the ALP-group is slightly higher than that between the two groups. In a few cases, however, the correlation between texts from different groups is even higher than
that between texts within the same group. For example, \( alp3 \) and \( elp3 \) correlate at 0.58, whereas \( alp3 \) and \( alp2 \) correlate only at 0.50.

There are relatively small differences among the correlations in Table 1, and there is no easily interpretable pattern. This is an indication that the eight texts measure on a single dimension, at least in the factor-analytic sense of the term. This interpretation is corroborated by the results of a factor analysis (maximum-likelihood method with varimax rotation), which yields only one substantial factor (eigenvalue = 4.47; 55.9% variance explained). In addition, there is a misfit between data and factor-analytic model \((p = 0.07)\).

With regard to the interpretation of the results of the factor analysis the following problem should be taken into account (see Grotjahn, 1987; 1992a and the references given therein): the results of factor analyses depend to a considerable extent on the degree of heterogeneity of the sample under investigation. If the subjects are very heterogeneous with regard to the measured ability, factor analysis tends to yield a substantial first factor.

In the present case for all eight texts the minimum score is 21, the maximum 138, the mean 91.7 and the standard deviation 21.9. There is thus a considerable degree of heterogeneity.

Therefore, in order to reduce the variance (heterogeneity), we have divided the sample into two ability groups. Since an independent measure of language proficiency was not available, we used the median of the C-test score \((\text{Med} = 91)\) as a cut-off point. For these relatively homogeneous groups two separate maximum-likelihood factor analyses with subsequent varimax rotation were then calculated.

For the low-ability group we now obtain two factors which explain 27.8% and 8.4% of the variance. For the high-ability group we also get two factors explaining 21.7% and 19.7% of the variance. In both groups the fit between model and data is good (low-ability group: \((\chi^2 = 7.4, df = 13, p = 0.88)\); high-ability group: \((\chi^2 = 11.1, df = 13, p = 0.60)\). The first factor is primarily due to the \( ALP \)-texts, whereas the second, weaker factor appears to measure mainly everyday language proficiency. Overall, this result indicates that the unifactorial solution in the total group may be attributed to the fact that differences between the type of text are obscured by differences in ability and that our data, aggregated into

<table>
<thead>
<tr>
<th></th>
<th>( elp1 )</th>
<th>( elp2 )</th>
<th>( elp3 )</th>
<th>( elp4 )</th>
<th>( alp1 )</th>
<th>( alp2 )</th>
<th>( alp3 )</th>
<th>( alp4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( elp1 )</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( elp2 )</td>
<td>0.60</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( elp3 )</td>
<td>0.57</td>
<td>0.53</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( elp4 )</td>
<td>0.59</td>
<td>0.54</td>
<td>0.72</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( alp1 )</td>
<td>0.51</td>
<td>0.51</td>
<td>0.61</td>
<td>0.61</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( alp2 )</td>
<td>0.53</td>
<td>0.52</td>
<td>0.62</td>
<td>0.60</td>
<td>0.60</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( alp3 )</td>
<td>0.47</td>
<td>0.48</td>
<td>0.58</td>
<td>0.55</td>
<td>0.55</td>
<td>0.50</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>( alp4 )</td>
<td>0.47</td>
<td>0.46</td>
<td>0.54</td>
<td>0.59</td>
<td>0.63</td>
<td>0.54</td>
<td>0.60</td>
<td>1.00</td>
</tr>
</tbody>
</table>
more homogeneous groups, provides evidence for separate everyday and academic language proficiencies.

If one uses oblique rotation (oblimin criterion), a clearer pattern of loadings emerges in both groups. However, in the low-ability group the two factors are correlated at 0.31, in the high-ability group even at 0.50. There is thus a considerable overlap between both dimensions. This confirms the model described above according to which the ability measured by the C-test consists of three major components: a specific everyday language proficiency, a specific academic language proficiency and general language proficiency (the latter corresponding to the correlation between the two factors).

The Classical Latent Additive Test Model

Another instrument for investigating the dimensionality of C-tests and other (language) tests is the Classical Latent Additive Test Model (CLA model). In the CLA model the true score of classical test theory is decomposed into a latent ability parameter (subject parameter) and a latent difficulty parameter (item parameter). It is assumed that the manifest test behaviour can be explained by an additive effect of the two latent parameters. If the model can be shown to be valid, the test is considered to measure unidimensionally and on an interval scale.

There are several possibilities to test the validity of the model empirically (see Grotjahn, 1987). A necessary, though not sufficient condition for validity is the additivity of the items (in the present case the C-test texts). Non-additivity means that there is an interaction between the ability of the subjects and the difficulty of the items. Additivity can be examined with the help of the Tukey test provided by SPSS within the procedure ‘reliability’.

Next the subject sample is divided into two samples according to some appropriate criterion (e.g. the median of the total score). The item difficulties are then estimated in both samples. If the CLA model is valid, the estimates have to be independent of the samples. This can be assumed if the observed differences between estimates from the two samples fail to reach statistical significance. Otherwise the corresponding item (i.e. text) can be eliminated. Especially problematic are texts with a difficulty estimate that is higher for the low-ability group than for the high-ability group. The necessary calculations can again be quite easily done with SPSS.

An inadequate model fit can also be due to some subjects reacting very differently from the rest of the group with regard to a specific text. The CLA model provides means to identify these subjects.

In the following we will not describe in detail the CLA analyses but rather present some important results. The reader interested in more specific information is referred to the analyses in Raatz (1985a, 1985b) and Grotjahn (1987, 1992a).

CLA analysis shows that the eight C-test texts are highly non-additive (Tukey test: \( p = 0.03 \)). Separate analyses for the C-test ELP and the C-test ALP also yield a strong effect of non-additivity. Particularly striking is the high negative \( t \)-value for the text elp2 \( (t = -5.18) \). This means that the text is more difficult for the high-ability group than for the low-ability group. This undesirable property of text elp2 may be due to certain lexical idiosyncrasies.

Additional calculations show that in the case of the scale ELP elimination of
elp1 and elp2 results in a good model fit without too much decreasing the reliability of the scale (for reliability see the next section). With regard to the scale ALP exclusion of texts alp1 and alp2 yields a similar result (exclusion of texts is here not absolutely necessary). If we now combine the texts elp3, elp4, alp3, alp4 into one single scale, the Tukey test of non-additivity is highly significant ($p = 0.007$). Thus, not only according to the factor analysis of homogeneous subgroups but also according to CLA analysis the scales ELP and ALP appear to tap different aspects of L2 ability.

**Reliability**

Calculation of Cronbach’s Alpha for all eight texts gives a value of 0.91. For the individual scales ELP and ALP the reliabilities are 0.85 and 0.84. Excluding the texts elp1, elp2, alp1 and alp2 from the scales yields 0.83 for ELP and 0.75 for ALP. Although the value for ALP has considerably decreased, it is still sufficiently high with regard to the objectives of the present study. Taking both the results of the CLA analysis and the reliability check into account we have decided to carry out the subsequent statistical analyses on the basis of the shortened scales ELP and ALP consisting of the texts elp3 and elp4, and alp3 and alp4 respectively.

Overall, the statistical analyses confirm hypothesis 1: it is possible to develop C-tests, which differentiate between everyday and academic language proficiency.

**Everyday and academic language proficiency**

Tables 2 and 3 show that compared to the control group the scores of the returnees are higher both in academic and in everyday language C-tests. However, a statistically significant difference can only be found in the everyday language tests. The results therefore corroborate hypothesis 2. The language proficiency of both groups differs in respect to the relation between academic and everyday language proficiency. The specific language acquisition process of the returnees leads to an advantage in everyday language proficiency but not in academic language proficiency. It has to be noted that the control group consists of a very small number of foreign language learners ($n = 10$) which limits the power of the tests of significance.

In Table 2, the high standard deviation in the group of returnees is particularly striking. This indicates that there is possibly a wide range of factors that influence

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Scores on the academic language C-test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
<td><strong>Mean</strong></td>
</tr>
<tr>
<td>Returnees</td>
<td>16.51</td>
</tr>
<tr>
<td>Control Group</td>
<td>14.90</td>
</tr>
<tr>
<td>Total</td>
<td>16.41</td>
</tr>
</tbody>
</table>

$t$-test, two-tailed, $p = 0.134$

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Scores of the everyday language C-test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
<td><strong>Mean</strong></td>
</tr>
<tr>
<td>Returnees</td>
<td>28.00</td>
</tr>
<tr>
<td>Control Group</td>
<td>20.40</td>
</tr>
<tr>
<td>Total</td>
<td>27.55</td>
</tr>
</tbody>
</table>

$t$-test, two-tailed, $p = 0.001$
the proficiency of the returnees, whereas the control group seems to be much more homogeneous.

It is important to bear in mind that the possible maximum score in both tests is 40 (two C-tests of 20 blanks). In relation to this maximum, the difference of 7.6 between the groups in Table 3 is therefore equivalent to 19%. Overall, the results indicate that the two groups have a different type of language proficiency. The returnees have a clear advantage in the everyday language proficiency whereas both groups have roughly equal scores at the academic language level. In order to illustrate this difference we computed the ratio between the two tests (scores C-test ELP : scores C-test ALP) as shown in Table 4.

Table 4 Ratio scores C-test ELP: scores C-test ALP

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Returnees</td>
<td>0.60</td>
<td>0.21</td>
<td>158*</td>
</tr>
<tr>
<td>Control Group</td>
<td>0.80</td>
<td>0.29</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>0.61</td>
<td>0.22</td>
<td>168</td>
</tr>
</tbody>
</table>

Mann-Whitney U-test, two-tailed, \( p = 0.015 \). For methodological reasons the non-parametric Mann-Whitney U-test is more appropriate than the t-test for this computation. *We could not compute the ratio for one person due to missing data.

There are two ways of interpreting the results obtained so far:

1. Returnees have a higher language proficiency in German than the control group, but this holds only for everyday language use (ELP).
2. The foreign language learners have a lower ELP in German than the returnees, but their ALP is quite high compared to the control group.

This means that the language proficiency profile of the returnees is characterised by a discrepancy between ELP and ALP (hypothesis 2).

Biographical data

In this section we try to determine which factors mainly influence the language proficiency as measured by our C-tests. We take the following predictor variables into consideration:

- duration of stay since return;
- occupation of the parents;
- sex;
- age;
- years of school attendance in Germany (school years);
- type of school attended in Germany (school type).

Tables 5, 6 and 7 show some statistics of the predictor variables.

The intercorrelations between the predictor variables are shown in Table 8. The highest correlation is between ‘school years’ (in Germany) and ‘duration of stay since return’. Those informants who came back earlier had less time to attend school in Germany. Therefore this correlation is negative (-0.72). The second largest correlation is between ‘school type’ and ‘school years’ (in Germany). This is logical because some school types (especially Gymnasium) go
up to year 13, whereas pupils in other school types leave school normally after year 10 (Hauptschule, Realschule).

One of our main questions in this study is whether the language proficiency at the everyday and academic level can be predicted on the basis of the language acquisition biography. Table 8 shows that there is a significant positive correlation between the scores in the two C-tests and ‘school years’ in Germany and ‘school type’ attended there, whereas there is a significant negative correlation between ‘duration of stay since return’ and the C-test scores. This means that students, who were longer at schools in Germany and attended a higher type of school, have a higher language proficiency as measured in both tests. Students who returned earlier to Turkey have lower scores in both tests. In order to investigate the influence of the predictor variables in more detail we carried out a multiple regression analysis with the variables in Table 8. The dependent variables were the scores on the two C-tests. We did not use more complex methods

### Table 5 Statistics of the variables ‘age’, ‘duration of stay since return’ and ‘school years’

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>20.3</td>
<td>17</td>
<td>28</td>
<td>1.9</td>
<td>159</td>
</tr>
<tr>
<td>Duration of stay since return</td>
<td>7.0</td>
<td>0</td>
<td>18</td>
<td>2.5</td>
<td>158</td>
</tr>
<tr>
<td>School years</td>
<td>6.8</td>
<td>0</td>
<td>13</td>
<td>2.2</td>
<td>156</td>
</tr>
</tbody>
</table>

### Table 6 Statistics of the variables ‘sex’ and ‘occupation of parents’

<table>
<thead>
<tr>
<th>Variable</th>
<th>Male</th>
<th>Female</th>
<th>No statement</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td>159</td>
</tr>
<tr>
<td>male:</td>
<td>92 (= 58%)</td>
<td>67 (= 42%)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Occupation of parents</td>
<td></td>
<td></td>
<td></td>
<td>159</td>
</tr>
<tr>
<td>unskilled workers:</td>
<td>123 (= 77%)</td>
<td>34 (= 21%)</td>
<td>2 (= 1%)</td>
<td></td>
</tr>
</tbody>
</table>

* skilled workers, employees and self-employed.

### Table 7 Statistics of the variable ‘school type in Germany’

<table>
<thead>
<tr>
<th>Type of school</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grundschule* (primary school)</td>
<td>32</td>
<td>20</td>
</tr>
<tr>
<td>Hauptschule** (elementary school)</td>
<td>56</td>
<td>35</td>
</tr>
<tr>
<td>Realschule (secondary school finishing with GCSEs)</td>
<td>46</td>
<td>29</td>
</tr>
<tr>
<td>Gymnasium (grammar school)</td>
<td>21</td>
<td>13</td>
</tr>
<tr>
<td>Gesamtschule (comprehensive school)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>No statement</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

* All informants in this category left Germany before secondary school.
** Originally the ‘Hauptschule’ was meant to be the main school for the preparation for non-academic occupations. In the last decades it has lost much of its reputation and has become what is sometime pejoratively called a ‘Restschule’, which means that everybody tries to get to other schools and only the pupils with poor academic performance (the ‘rest’) stay at the ‘Hauptschule’.
such as LISREL analysis because the sample is relatively small. For the same reason we reduced the possible values of the variable ‘school type’ to two: ‘Grund- und Hauptschule’ (primary and elementary secondary) and ‘all other types’. We double-checked all results by carrying out all computations with three different procedures: stepwise selection, forward selection and backward elimination. In all cases the SPSS default values were used: $p = 0.05$ for selection and $p = 0.10$ for elimination of variables. Furthermore we investigated whether there were violations of the general assumptions on which regression analysis is based (see also Grotjahn, 1992b). Although there are some violations mainly due to the intercorrelations between predictor variables (see Table 8), all violations are within acceptable bounds.

Different procedures, namely stepwise selection and forward selection, produce identical results. The accumulated variance explained by the predictor variables is displayed in Table 9. The determination coefficients are adjusted according to the number of predictor variables.

Table 9 shows that predictor variables related to school attendance in Germany (‘school years’ and ‘school type’) appear to have the strongest influence on language proficiency in German. This holds true for ALP as well as ELP. School years and school type jointly explain 22% of the variance of ALP and 24% of ELP. The longer the school attendance and the higher the school level the better are the results in both proficiency domains. Note, however, that the order of the two factors is reversed: in the case of ALP ‘school years’ is the most important

### Table 8 Correlations between predictor variables and dependent variables

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Duration</th>
<th>School Years</th>
<th>Sex</th>
<th>Occupation</th>
<th>School Type</th>
<th>ELP</th>
<th>ALP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td>0.35**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School years</td>
<td>0.22**</td>
<td>-0.72**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>0.34**</td>
<td>0.18</td>
<td>-0.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td>0.02</td>
<td>-0.09</td>
<td>0.14</td>
<td>-0.13</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School type</td>
<td>-0.06</td>
<td>-0.37**</td>
<td>-0.47**</td>
<td>-0.10</td>
<td>0.10</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELP</td>
<td>0.08</td>
<td>-0.32**</td>
<td>0.41**</td>
<td>-0.02</td>
<td>0.24*</td>
<td>0.43**</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>ALP</td>
<td>0.14</td>
<td>-0.27**</td>
<td>0.43**</td>
<td>-0.12</td>
<td>0.20*</td>
<td>0.37**</td>
<td>0.69**</td>
<td>1.00</td>
</tr>
</tbody>
</table>

‘Duration’ = duration of stay since return; ‘school years’ = school years in Germany; ‘occupation’ = occupation of parents; ‘school type’ = school type in Germany; ‘ELP’ = scores C-test ELP; ‘ALP’ = scores C-test ALP.

* $p < 0.010$ (one-tailed); ** $p < 0.001$ (one-tailed)

<table>
<thead>
<tr>
<th></th>
<th>Academic Language Proficiency</th>
<th>Everyday Language Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>School years</td>
<td>0.18</td>
<td>School type</td>
</tr>
<tr>
<td>School type</td>
<td>0.22</td>
<td>School years</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Occupation of parents</td>
</tr>
</tbody>
</table>
factor, whereas in the case of ELP ‘school type’ is most important. For ELP ‘occupation of parents’ is an additional but minor factor explaining only 3% of the variance. For this reason it seems reasonable to restrict the interpretation to ‘school years’ and ‘school type’.9

If one runs in addition to stepwise and forward selection the procedure backward elimination, one obtains somewhat different results. In the case of ALP the two school-related variables remain the main predictor variables with an additional third variable, namely ‘occupation of parents’, increasing the explained variance to 24%. While the method ‘backward elimination’ thus yields roughly the same picture for ALP it produces different results for ELP. The variable ‘school years’ remains the main predictor variable. However, in addition the variable ‘duration of stay since return’ plays a major role. This variable shows a negative beta value (beta = -0.25). This means that ‘duration of stay since return’ to Turkey affects the everyday language proficiency of German negatively. This is a clear indication of language attrition in the domain of ELP. For the interpretation of this difference between ELP and ALP one has to bear in mind that in our sample the returnees have maintained contact with the German language since return mainly in an academic environment (school and university). In addition to the variables ‘school years’ and ‘duration of stay since return’ backward elimination yields a third variable for ELP, namely ‘age’. Older returnees generally have lower everyday language proficiency than the younger ones. A clear-cut interpretation of the influence of age is however difficult because of the high intercorrelations between the age of the informants and the variables ‘school years’ and ‘duration of stay since return’.

Conclusions

The statistical analysis suggests a model of language proficiency that consists of at least two main components: a core area, which could be labelled as general language proficiency and a more task-specific area. In our case the latter consists of everyday language and academic language proficiency. Further research has to show whether general and everyday language proficiency as measured by the C-test are identical to a large extent. The results of the present study support the proficiency model of Jim Cummins. C-tests on the basis of different types of text are one possible way of operationalising Cummins’ framework for language proficiency. Different types of C-tests can be developed as a useful means of analysing different aspects of language proficiency.

Overall, the present study shows that compared to the control group there is a large gap between the level of ELP and ALP of the returnees. Returnees score significantly higher than the control group on C-test ELP but not on C-test ALP despite the fact that the returnees spent several years in Germany and at German schools whereas the control group had contact with German only in a classroom setting. This means that in this specific migration situation the returnees were able to turn the long and intensive contact with the target language to their advantage only with regard to ELP.

The main factors which influence the level of language proficiency at the everyday and academic level are ‘years of school’ and ‘type of school’ in Germany and are thus both school related. These school variables are still impor-
tant even seven years after return. All methods of statistical analysis identify these two factors. The method ‘backward elimination’ yields slightly different results. It does not contradict the other methods but gives an additional piece of information: it produces ‘duration of stay since return’ as an additional factor. A long duration of stay in Turkey after return affects the scores of the returnees in the everyday C-test negatively. There are two possible interpretations: (a) language attrition after return; and (b) insufficient language acquisition of German before return.

Both factors will probably be of some influence. However, we tend to attribute a more important role to factor (a) for the following reasons: everyday language proficiency is acquired in a few years, whereas the development of academic language proficiency needs much more time (see Theoretical Background above). We therefore assume that even those subjects who returned very young had a certain proficiency at everyday level. A lower level of everyday language proficiency for returnees who had a long duration of stay after return because they came back earlier is therefore more likely to be due to language attrition. There are no indications that the duration of stay affects the academic level. The reason for this is probably that after their return the returnees stayed in contact with academic German either at school or university.

The main characteristics of the language proficiency of the returnees in our sample are the following. There is a considerable difference between ELP and ALP, the former being much higher with returnees than with foreign language learners. However, once returned to Turkey a process of language attrition takes place in the domain of ELP. A possible similar process in the domain of ALP is compensated by contact with academic German in classroom and at university. As a result, while ELP decreases, ALP is maintained or even increases. Overall this means that for the returnees there is a shift towards a proficiency profile that is more typical of foreign language learners than for bilinguals who learned German as a foreign language.

Acknowledgements

We would like to thank Eoghan Mac Aogáin, Kate Beeching and Volker Green for comments on an earlier version of this article.

Correspondence

Any correspondence should be directed to Dr Helmut Daller, Faculty of Languages and European Studies, University of the West of England, Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY, UK (Helmut.Daller@uwe.ac.uk).

Notes

1. In 1994 there were about 5000 university students of German in Turkey (Yıldız, 1994). An unpublished survey carried out in 1992 by Helmut Daller showed that at least 95% of these students were returnees.
2. We distinguish between everyday and academic language proficiency. The relation between this distinction and Cummins’ notion of BICS versus CALP is discussed in more detail in the next section. Closely related and established terms are primary versus secondary language skills (see Baur & Meder, 1992).
3. The C-test is a variant of the cloze test. Starting with the second sentence of a text every
second half of every second word is deleted. A C-test normally consists of 4–6 texts with 20 blanks per text (for a general overview see Grotjahn, 1995).

4. Arguably there is no such thing as a ‘neutral’ text. In this article we cannot go into more detail about this point.

5. The CLA model has already been used e.g. by Raatz (1985a, 1985b) and Grotjahn (1987) to examine the dimensionality of C-tests. The model is dealt with rather extensively in Steyer and Eid (1993).

6. Other methods for the analysis of possible interactions between subject ability and item difficulty are presented e.g. in Holland and Weiner (1993). More comprehensive analyses must of course take into account this aspect as well.

7. These findings match similar results of earlier research with returnee children (see Daller, 1995).

8. In a preliminary analysis we also included ‘length of residence in Germany’. We excluded this factor, however, because of its high intercorrelation with ‘duration of stay since return’.

9. The variables ‘school years’ and ‘duration of stay since return’ correlate at $r = -0.72$ (see Table 8). They thus measure to a considerable extent the same trait. In spite of this fact and the ensuing methodological problems, we have decided to use both variables in the regression analysis, since we feel that each variable can make a specific contribution to the prediction of language proficiency (‘duration of stay since return’, for example, may be related to language attrition).

References


Appendix

C-Test Elp

elp1

Toni Schumacher


elp2

Türkische Kaffehäuser

Immer in Fahrt


Wer ist St. Nikolaus?


C-Test ALP

Literatursprache


Normen in der Grammatik

Die Junggrammatiker


Theoriebildung in der Soziolinguistik