Research Article

Audrey Bonvin*, Ladina Brugger and Raphael Berthele

Lexical measures as a proxy for bilingual language dominance?

https://doi.org/10.1515/iral-2020-0093
Received June 30, 2020; accepted May 3, 2021; published online May 21, 2021

Abstract: In bilingualism research, language dominance has been conceptualized and operationalized in different ways. In this paper, we discuss this notion and investigate to what extent language dominance is congruent with vocabulary knowledge in each language. First, we summarize definitions and operationalizations of language dominance. We show that comparison of proficiency related measures is a common operationalization of language dominance. Yet, the importance of attitudinal, biographical, or use-related components is often stressed. Such components are included in survey instruments like the Bilingual Language Profile (BLP). Second, we analyze data on language profiles of 225 French/German and 70 Italian/German adult bilinguals. Correlation and regression models are fitted to investigate the relation between the multi-dimensional dominance metric (BLP) results and results based on lexical tests (LexTALE for German and French, the VSPT from Dialang for Italian). The results reveal a strong linear association between BLP and LexTale.

Keywords: adult bilinguals; bilingual language profile; language dominance; LexTALE; vocabulary

*Corresponding author: Audrey Bonvin, Multilingualism and Foreign Language Education, Faculty of Arts and Humanities, University of Fribourg, Fribourg, Switzerland, E-mail: audrey.bonvin2@unifr.ch. https://orcid.org/0000-0003-4132-2391
Ladina Brugger, Bern University of Teacher Education, Bern, Switzerland, E-mail: Ladina.Brugger@phbern.ch. https://orcid.org/0000-0001-8860-3540
Raphael Berthele, Institute of Multilingualism, University of Fribourg and University of Teacher Education of Fribourg, Fribourg, Switzerland, E-mail: raphael.berthele@unifr.ch. https://orcid.org/0000-0003-1403-4116

Open Access. © 2021 Audrey Bonvin et al., published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.
1 Introduction

The notion of dominance is well-established in bilingualism literature. It is most commonly seen as being distinct and distinguishable from skills or proficiency in the languages of the individual: “We distinguish bilingual groups with respect to their proficiency in the L2, their relative language dominance, and the degree to which the context of language use supports each of the two languages.” (Kroll and Linck 2007: 238).

As this illustrates, dominance as a construct is generally used as a grouping or continuous variable, with the typical goal of explaining variance in an outcome variable that operationalizes an aspect of language skill or competence. It can also be used as an outcome variable in that researchers may want to find out what the dominant language is in certain bi- or multilingual individuals as in Bahrick et al. (1994) or Zirkel (1974). A third perspective on dominance is the investigation of the notion itself as a compound construct, with a focus on its internal dimensionality. This third perspective is usually chosen in relation to the first – that is, scholars seek the best metrics in the service of the prediction of outcome variables (e.g. Dunn and Fox Tree 2009; Gertken et al. 2014). Different composite measures of dominance have been proposed, and we use one of them, the Bilingual Language Profile, here (Birdsong et al. 2012).

We first give an overview of how the notion of dominance is used and operationalized in bilingualism literature. We then investigate, drawing on data from 295 adult bilinguals, the extent to which a composite measure of dominance is linearly associated with scores on tests of receptive vocabulary knowledge. The final part of the article discusses to what extent and under which conditions one measure might stand in for the other, i.e. when tests of vocabulary knowledge might be just as good as dominance measures and vice versa.

1.1 Constructs and operationalizations of language dominance

Although the term dominance is widely used in the literature, its denotation in the field is polysemous to say the least. Dominance can refer to a relative ‘importance’ or ‘weight’ in bilinguals’ repertoires without much further explanation (see the quote by Kroll and Linck above) but with a clear distinction made between skill and use on the one hand and dominance on the other. Other authors use dominance to refer to “the observed asymmetries of skill in, or use of, one language over the other” (Birdsong 2014: 374). Whereas for some dominance is different from or additional to relative skills and use patterns, for many other authors it roughly

Another approach is to operationalize language dominance in terms of reported or – more rarely – observed use of languages at home and in the workplace, as e.g. in Sebastián-Gallés et al. (2005: 253) (language use at home) or in Bedore et al. (2012) (current use). Other use-oriented metrics can be found, like language use for internal thought (Rose 1980). Sometimes use and proficiency are taken together, as e.g. in Pease-Alvarez and Hakuta (1993).

A yet different take on language dominance draws on language history, often with a particular focus on a person’s age at the beginning of the acquisition/learning of the second or additional language, or on an estimation of the cumulated amount of exposure (Unsworth 2015, see Birdong 2014 for a critical discussion of the need to disentangle age and dominance).

Finally, attitudinal measures have also been suggested for the operationalization of dominance, as e.g. a simple question as to which language would be kept if the participants had to choose (Cutler et al. 1989). Amengual (2018: 14) also argues in favor of attitudes as “an important aspect of dominance”.

Several of these studies focus on children and therefore need tools of elicitation other than the questionnaire-based methods discussed in this article. Nevertheless, the list of manifest variables presented here gives an impression of the variety of measures that arguably tap into language dominance, and this list can easily be extended. This diversity of operationalizations shows that, as it is the case for other constructs used in linguistics or psychology, there is no agreement whatsoever as to what a good measure of language dominance is. It also indicates a lack of consensus about whether and how dominance can be conceptually and empirically separated from skills, current use or past exposure, or attitudes.

Therefore, some scholars have argued that dominance is a compound construct that draws on diverse conceptual dimensions (skills) and domains (contexts of language use) (cf. Birdsong 2015). Examples of such metrics are the Bilingual Dominance Scale BDS (Dunn and Fox Tree 2009) and the Bilingual Language Profile BLP (Birdsong et al. 2012).
Nevertheless, as rightfully pointed out by a reviewer of this paper, the search for a universal definition and a single all-purpose operationalization of language dominance is unrealistic, partly because of the complexity and the dynamism of the construct. In this view, language dominance is “task-specific and domain-specific” and what is important is to clearly define the construct, explain its operationalization and the dimensions that are measured in each study (cf. Treffers-Daller 2015: 265). While we agree with the call for context-sensitive nuance in this domain (see end of Section 1.2), we still believe it is worth trying to find a metric for the core construct based on converging evidence from different contexts. Validation studies can help finding points of agreement on metric(s), which could be used to compare results between studies. Moreover, based on a growing sample of studies using general measures such as the BLP, it will be possible to capture its association (or that of its subcomponents) with language behavior in bilinguals.

The BDS is a questionnaire that contains questions on the percentage of language use, on the age of onset and degree of ‘comfort’ in a language, and on fluency. Its validity was tested by assessing the linear association of the numerical results obtained via the questionnaire with behavioral measures such as scores on a translation task, reaction times, filler rates and elongation rates in spoken use of the languages (Number of participants = 55; Dunn and Fox Tree 2009: 284).

The BLP is another questionnaire containing questions on more dimensions and domains than the BDS. It contains several questions on language proficiency, use, history, and attitudes regarding both languages of the individual (see Section 2.1.1). The instrument was developed using factor analytical techniques that support the internal dimensionality of the instrument. The authors use tests of association with proficiency tests and other behavioral measures to claim the validity and reliability of their instrument, but also to argue that the moderate association of the BLP instrument and, say, measures of processing (e.g. in naming tasks), shows that the BLP does not measure exactly the same construct as these behavioral tasks (Number of participants = 68; Gertken et al. 2014: 218).

Solís-Barroso and Stefanich (2019) compared four measures of language dominance: The BLP, the BDS, a self-rating of verbal/written ability and a repetition task. Among others, they found a positive association ($r = 0.69$) between the BLP and the BDS in a cohort of 29 bilinguals.

1.2 Conceptual variability of dominance

The notion of dominance is immediately appealing not only to bi- and multilingualism scholars, but also to bilinguals themselves: Languages in the speakers’
reperitories have different weights and play different roles, and the relative weight

can shift from one to the other depending on the bilingual’s biographies (cf. the

language attrition literature discussed in Montrul 2016) or use domains. In this

section we discuss conceptual and practical problems with composite dominance

metrics such as the BLP or the BDS. These problems are the backdrop of our

analysis of the association of the BLP with the uni-dimensional construct of an

individual’s vocabulary size.

Compound measures based on questionnaire data, just like less complex

operationalizations of dominance, are systematically but not perfectly associated

with scores on behavioral tests of various language skills and competences. The

claim that dominance is or is not conceptually separable from such behavioral

measures and the constructs they represent is thus axiomatic. Since so far no

completely independent way of testing language dominance has been suggested,
circularity is unavoidable: Interview or questionnaire-based dominance indices

can only be validated by testing their association with other behavioral measures.

If the analysis shows that they are not associated at all, that would be theoretically

and empirically challenging for the whole field of bilingualism research. If they

turn out to be moderately associated (which they usually are), scholars may claim,
as in the case cited above, that they are ‘valid’ but that the considerable amount of

residuals of the linear association shows that the two instruments do not measure

exactly the same constructs. If they were almost perfectly associated (e.g. with $r$

coefficients around 0.9), then the question arises to what extent the construct of

dominance can and should be separated from the construct tested by the other

behavioral task. Therefore, it is unsurprising if dominance indices obtained with

instruments such as the BDS or BLP are ‘significant’ predictors of an outcome

variable (skills, reaction times, etc.), since the definition of dominance is at least

partially reliant on language proficiency.

Questionnaires such as the BLP sometimes need to be adapted to specific

language ecologies. Not always can both languages in the repertoire be used in the

same contexts, for example in education. As argued in Poarch et al. (2019), this

applies e.g. for bidialectals, since in cases of diglossia with bilingualism (Fishman

1967) very often no formal education takes place in one of the two varieties. Items

such as the question “How many years of classes (grammar, history, math, etc.)

have you had in the following languages (primary school through university)?”

(BLP) therefore need to be adapted or dropped for contexts of bilingualism

involving dialects or languages that are not generally used for literacy.

Moreover, since such measures provide a subtraction-derived index, the

problem of being balanced on high or low levels of proficiency is addressed by
some authors (e.g. Birdsong 2015; Gertken et al. 2014; Treffers-Daller 2015: 260–262). An alternative is the use of a ratio-derived index instead of a subtraction. In both cases, however, “useful information concerning raw performance in each language is lost” (Birdsong 2015: 87–91). Besides, bilinguals with an identical dominance index usually differ for their answers to individual items (Birdsong 2015: 100). Another issue is certainly that the validity and reliability, on top of the conceptual circularity mentioned above, were evaluated with small samples and a limited set of languages.

Also, the decisions on how much weight the different dimensions should have for the total dominance indices seem fairly arbitrary, at least at first glance: The BLP, for example, gives equal weight to the four components (proficiency, use, history, and attitudes). A rationale behind this decision is provided by the following citation from Gertken et al. (2014: 215):

> Unlike Dunn and Fox Tree (2009) we decided to not differentially weight the four components of the BLP. To weight, say, the scores on the proficiency module above those on the other modules would bias the global scores for dominance and, effectively, our operationalization of dominance and users’ interpretation of dominance toward reflecting proficiency at the expense of the other components. This said, those users who wish to concentrate on proficiency assessment are able to do so – either independently of, or relative to, scores on the other components (There are instructions on the BLP website for this purpose; Birdsong et al. 2012)

According to the authors, the most correct account of language dominance as defined holistically consist of giving each component the same weight. This scoring practice also prevents minimizing “the risk of inflating the significance of any one factor” (section “Feature Comparison” on the website, Birdsong et al. 2012). Indeed, this may be questionable in terms of empirical evidence for generally giving more weight to one or some component(s). In this view, we understand neutral weighting as a compromise to propose an instrument “relevant for diverse bilingual populations” (Gertken et al. 2014: 215).

Overall, the notion has success in the literature but work on improving both conceptual clarity and empirical foundation is still necessary (cf. Treffers-Daller 2015). We think, as most scholars in bilingualism seem to do, that the notion is intuitively appealing and captures something that bilinguals experience: Bilinguals’ languages (and dialects) have different roles and a different status respectively, and this can also be expressed in terms of relative or differential weight or balance. One way of addressing this intuition is to allow for varying and multi-dimensional dominance: One language could be dominant affectively, but

---

1 We use Birdsong (2015: 86) terminology of (ratio-derived or subtraction-derived) index to refer to the results of a mathematic formula comparing two languages and score to report results for each language independently.
not in terms of use or proficiency, etc.; or dominance may even be construed as being temporally unstable, with shifting dominance relationships across contexts, situations, and, of course, across the lifespan.

In the remainder of this contribution, we attempt to gain further insight into the notion of language dominance. We ask to what extent the compound operationalization as instantiated in the BLP provides the same classification of participants on a continuum as a receptive vocabulary test. This basic question has practical relevance for research, since, under time or resource constraints, collecting one type of data may be easier than the other. Our contribution is therefore geared towards the needs of the bilingualism research community.

2 The study

To answer the question of what the respective information is in the BLP and vocabulary tests, we draw on data about the language profile of bilingual participants in Switzerland. We analyze two language pairs: French and German as well as Italian and German. The French/German bilinguals data comes from two projects. Project FR&DE1 is about information structures in narratives (ongoing Ph.D. project of the first author) and project FR&DE2 was on the use of motion descriptions (Brugger 2017; Stocker and Berthele 2020). As a comparison, we analyze the data of project IT&DE, which is about information structure in Italian/German bilinguals (IT&DE, ongoing Ph.D. project by Giulia Berchio). In all three projects, researchers administered the BLP survey as well as a receptive vocabulary test in the two main languages of the participants. In this paper, we focus on the results obtained by these instruments.

2.1 Descriptions of the research instruments

2.1.1 The Bilingual Language Profile (BLP)

This measure of language dominance contains an introductory section for biographic information and four modules including questions about language history, language use, language proficiency, and language attitudes. Each of its 19 questions is asked for both languages.

---

2 This section was adapted from Brugger (2017: 84–86).
Since participants spoke Swiss German and, consequently, standard German at a high level as well as other languages, we asked them to consider standard German as well as dialectal varieties as one language to answer the BLP. For the purposes of the present study, we use the term “German” interchangeably to refer to the standardized variety or the dialects. If the specific variant is relevant to our argument, we mention it explicitly.

The first module on language history contains six questions. The first elicits the age of language learning onset, and the second asks the age at which a participant felt comfortable using the language. The next four items concern the number of years pursuing school lectures in each of the two languages, residing in a region/country where the languages are spoken, living with a family (household in our versions) and working in an environment where the languages were spoken.

The second module aims to gather information on language use through five questions. The first three questions are about the average percentage of use of French, German as well as other languages in an average week with friends, family (respectively household) and at work or school. The fourth and fifth questions elicit how often the participant thinks and counts in German and French.

In the third module, on language proficiency, four questions on abilities in speaking, understanding, writing, and reading are asked. Participants self-evaluate their abilities on a 7-point Likert scale.

The last module on language attitudes comprises four questions on a 7-point Likert scale. An example of a question is about the degree to which candidates feel like themselves when speaking each language.

As suggested by the authors of the BLP, their open-source instrument can be adapted to specific linguistic situations. Some minor modifications from the original questionnaire were made in all our three projects, to increase the validity of the participants’ responses by fitting the questions to the local context. Additional questions were asked but are not analyzed. These changes are described in the online appendices: https://osf.io/vqe46/?view_only=61f5a453e1ee4629acf0fa52411af0b2.

2.1.2 Lexical test for advanced learners of French and German (LexTALE)

The measure of receptive vocabulary knowledge to be compared with the BLP measure of language dominance is a lexical decision task, which contains about

---

3 In formal contexts in Switzerland, standard German is used predominantly and the dialects are mostly used in informal contexts. There are different approaches to describing this language situation in German-speaking Switzerland (Berthele 2014) and attitudes towards the different varieties are complex (Studler 2017).
one-third non-words and two-thirds words. LexTALE has been developed for the needs of psycholinguists working on advanced L2 speakers (Lemhöfer and Broersma 2012). It was first created for English, but versions for German (www.lextale.com) and French (Brysbaert 2013) now exist. In this paper, we use the terms LexTALE_DE for the test in German and LexTALE_FR for the test in French. Although they are structurally similar, there are some fundamental differences. For example, LexTALE_FR included more words to target participants on a larger proficiency range: “from people without (much) knowledge of French to advanced native speakers” (Brysbaert 2013: 31–32). For these and other reasons, the different lexical tests cannot be mapped onto a common, language-independent scale of lexical proficiency (see also Treffers-Daller 2015: 257). What can be derived from these metrics, nonetheless, are differences between participants within the language tested.

Lemhöfer and Broersma (2012) ran a validation study with 72 Dutch and 87 Korean learners of English. They first observed positive correlations between the scores of LexTALE and diverse translation tests ($r$ between 0.66 and 0.78 for Dutch participants, between 0.46 and 0.51 for Korean participants). Second, LexTALE scores were positively correlated with the Quick Placement Test, a general measure of language proficiency ($r = 0.63$ for Dutch participants and 0.29 for Korean participants). They concluded that LexTALE is “a good predictor” of general English lexical knowledge and not only receptive knowledge (Lemhöfer and Broersma 2012: 1). Currently, we are not aware of validation studies for the German and French versions of LexTALE. Brysbaert (2013: 33) reports a high reliability of LexTALE_FR in his study (Cronbach’s alpha of 0.96).

### 2.1.3 The vocabulary size placement test (VSPT) from Dialang

As no Italian version of LexTALE was available at the time of data collection, another word/non-word recognition test was used in project IT&DE. The VSPT from Dialang is based on verbs only and is part of a broader assessment process available to people pursuing further language learning. It is used to quickly obtain a global measure that allows the system to adapt the subsequent test items to match the proficiency level of the test-takers (Huhta 2007). For the sake of our research here, what matters is that the tests used are structurally equivalent and tap into the same type of competence.

---

4 Meanwhile, LexITA was published for Italian (Amenta et al., 2020).
Tests that measure reliably the same level of proficiency in several languages are rare since the vocabulary of the different languages is different in size and internal structure. As far as we know, there are no standardized vocabulary tests for adults that are anchored to the same, language-neutral proficiency scale.

2.2 Research questions and hypothesis

**Question 1.** What is the strength and direction of the relation between information obtained via a multi-dimensional dominance survey (BLP) and information obtained via a vocabulary recognition test (LexTALE)?

**Hypothesis:** Based on the insights from the literature review, we expect a strong positive relationship between the results from the BLP and the results from the LexTALE_DE/LexTALE_FR. However, age at testing could arguably influence both results: Changes over the lifespan in crystallized intelligence could lead to proportionally better results in LexTALE for older participants (e.g., Vanhove and Berthele 2015). The language dominance of a person may shift (several times) during life without necessarily implying a comparable shift in his/her language proficiency (Birdsong 2014: 378–379). Thus, we also check whether the strength of the relation between the BLP and the LexTALE changes across ages.

**Question 2.** How are the components of the dominance construct – language history, use, proficiency, and attitudes – related to the vocabulary test LexTALE?

**Hypothesis regarding language proficiency only:** Vocabulary knowledge as tested by lexTALE is central for all language skills that are measured by the proficiency module of the BLP (reading, writing, speaking, and listening). As for the method of self-assessment, Brysbaert (2013) found a correlation of 0.58 between the LexTALE_FR and self-assessment from 310 participants. Our data are heterogeneous regarding age and education level, which may increase the variation in self-assessed proficiency for the same LexTALE score (cf. Lemhöfer and Broersma 2012). Therefore, we expect a positive and medium relationship between the self-evaluation of proficiency of the BLP and the LexTALE for each language.

**Hypothesis taking all components of the BLP into account:** We expect language proficiency to be the strongest predictor of the LexTALE result, followed by language history and language use or vice-versa. We do not have *a priori* expectations regarding language attitude.

**Question 3.** Is the relationship between the BLP and the lexical decision task in French/German bilinguals mirrored in data from Italian/German bilinguals completing the Dialang test instead of LexTALE for Italian?
Hypothesis: We hypothesize a positive relationship between language dominance measure and lexical decision task. As Dialang is a placement test, there is the risk of obtaining a ceiling effect with participants highly proficient in Italian. This may contribute to a weaker relation between the two indices than in the French/German bilingual sample.

3 Method

3.1 Participants

The majority of the participants in projects FR&DE1 and FR&DE2 lived in or close to the bilingual town of Fribourg. Participants from project IT&DE lived in different cantons of German- or Italian-speaking Switzerland. These cohorts contain early and late bilinguals (Figure 1).

<table>
<thead>
<tr>
<th>Projects</th>
<th>FR&amp;DE2</th>
<th>FR&amp;DE1</th>
<th>IT&amp;DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants number</td>
<td>154 (1069)</td>
<td>71 (539)</td>
<td>70 (449)</td>
</tr>
<tr>
<td>Age</td>
<td>mean: 24, median: 23, range 18-48</td>
<td>mean: 37, median: 27, range 18-69</td>
<td>mean: 46, median: 48, range 18-85</td>
</tr>
<tr>
<td>Level of education</td>
<td>University students</td>
<td>Mixed</td>
<td>Mixed</td>
</tr>
<tr>
<td>Main languages</td>
<td>(Swiss) German and French</td>
<td>(Swiss) German and Italian</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Participants profiles of the three projects. The differences in the data characteristics are due to the specific aims of each original research proposal.

3.2 Data collection

The BLP was administered online. Participants could choose to answer the survey in German or French (or Italian). It was the first activity they completed. The majority completed the survey before they met the researcher.

For the lexical decision tasks, instructions from the English LexTALE were used (Lemhöfer and Broersma 2012: 329). Approximately half of the participants first completed the test in German and the other half in their other language (French or Italian). The two lexical decision tests were administered a minimum of two weeks apart and in the presence of the researcher. In project FR&DE1 and IT&DE,
participants took the test online. They could only see one item at a time and could not come back after validating the answer. In project FR&DE2, LexTALE was administered on paper. In this case, participants saw all items during the whole test.

### 3.3 Data processing

The BLP indices were calculated following the scoring instructions on the BLP website (Birdsong et al. 2012). First, a score for each language separately was calculated, in which each category is given the same weight. Then, the individual scores for French (or Italian) were subtracted from the scores for German to obtain an index of language dominance. Indices are distributed on a scale between $-218$ (maximal score for French or Italian and no score for German) and $+218$ (maximal score for German and no score for French or Italian), where 0 represents balanced bilingualism.

For LexTALE, the scoring procedure “% correctav (averaged % correct)” advised by Lemhöfer and Broersma (2012: 330) was used and adapted to the French and Italian vocabulary tests. The first three items of the French and German LexTALE are training items and excluded from the final score. For Dialang the first three items were not considered as training items. As for the BLP, the individual vocabulary test scores of the French LexTALE version were subtracted from the German LexTALE scores. Subtraction-derived indices are distributed on a scale between $-100$ (a score of 100 in French and 0 in German) and $+100$ (a score of 100 in German and 0 in French, respectively Italian), where 0 means that the participant obtained the same score for both languages (see Figures 2 and 10).

This procedure allowed us to obtain two subtraction-derived indices per participant on continua. We consider 0 as a benchmark for a balanced result for practical reasons in answering questions 1 and 3. We do not mean that this precise index is the (only) one corresponding to (perfect) balanced bilingualism (for a summary of the discussion concerning the notion of balanced bilingualism and its operationalization, see Treffers-Daller 2015: 241–248 and Birdsong 2015: 95–97). In what follows, we report these indices as well as the scores in each language.

The two density plots in Figure 2 illustrate the distribution of the data in projects FR&DE1 and FR&DE2. Data from project IT&DE are presented in the result part of the third question (cf. Figure 10).

The distribution of the data around the zero-value in Figure 2a and b looks comparable. Let’s have a look at the numbers behind both figures. The BLP mean (10.5) and median (16.7) (index range: $-218; 218$) for both projects together suggest that participants tend to be slightly more dominant in German, whereas the mean of the LexTALE index (0.4) and its median (1.2, index range: $-100; 100$) are more centered toward zero.

The outlier on the right side of Figure 2b is a participant who achieved much better results in the LexTALE_DE (83.75/100 points) than in the LexTALE_FR (5/100 points) but estimated her level of proficiency in the two languages as equal.
in the BLP survey (both 23/24 points). Since we failed to detect any clearly satisfying explanation for this result, we kept this participant in the data for the analysis (unless explicitly stated otherwise below).

4 Results

Statistics were carried out with the software R (R Core Team 2020). They are based on non-standardized scores. For the analysis with the outlier, we report Spearman rank correlation, and for the analysis without it, Pearson’s product-moment correlation. The simple linear models were fitted with the lm() function and the linear mixed-effects models with the lmer() function in the lme4 package (Bates et al. 2015) for R. For more information on the data and on analyses please refer to the supplementary material online (https://osf.io/vqe46/?view_only=61f5a453e1ee4629acf0fa52411af0b2).

4.1 Question 1

What is the strength and direction of the relation between information obtained via a multi-dimensional dominance survey (BLP) and information obtained via a vocabulary recognition test (LexTALE)?
First, we analyze the BLP and LexTALE subtraction-derived indices. Then, we look at these indices across ages. Finally, we compare the BLP and LexTALE scores for each language separately.

4.1.1 The relation between BLP and LexTALE indices

If the information provided by the two tests is comparable, the data points should be on a sloped line and the crossing lines of the two 0-indices should be on the trend line. Figure 3 shows that this is quite the case, the crossing lines of the two 0-indices being in the confidence band of the trend line.

The Spearman correlations ($r_s$) between the LexTALE and BLP indices is 0.86, indicating a strong positive relation between the two measures.

![Figure 3](image.png)

**Figure 3:** The x-axis represents the subtraction-derived BLP index (which ranges from −218 to 218). The further to the left, the more the participant is considered dominant in French, and vice versa. The y-axis represents the subtraction-derived LexTALE index (which ranges from −100 to 100). The further up, the more the participant got a better result in German in comparison to French, and vice versa. Each point represents a participant. The black line summarizes the trend. The 95% confidence band is in grey.

To explore how adequately the BLP index is predictable by the LexTALE index only, we fitted a linear regression model to the data (1). Due to issues with the model assumptions (linearity, constant variance, and normality) caused by the outlier mentioned above we also fitted the regression without this participant (2) (e.g. Vanhove 2018, 2019):

\[
\text{Predicted BLP index} = 9.3 + (3.3 \pm 0.2) \times \text{LexTALE index} \\
(\beta \pm \text{Standard Error, } p < 0.01).
\]

Standard deviation = 42. Multiple $R^2 = 66\%$. 

### Footnotes:

1. With the outlier: \[ \text{Predicted BLP index} = 9.3 + (3.3 \pm 0.2) \times \text{LexTALE index} \] 
   \[ (\beta \pm \text{Standard Error, } p < 0.01). \]

2. Standard deviation = 42. Multiple $R^2 = 66\%$. 

(2) Without the outlier: \( \text{Predicted BLP index} = 10.8 + (3.7 \pm 0.1) \times \text{LexTALE index} \) \((p < 0.01)\).

Standard deviation = 35. Multiple \( R^2 = 77\% \).

The BLP index can be predicted to a relatively large extent by the LexTALE index. The model is influenced to a considerable degree by the outlier. While the big picture remains the same, the model quite unsurprisingly captures more variability in the BLP for the sample without the outlier (see \( R^2 \)).

### 4.1.2 The relation between the LexTALE and BLP subtraction-derived indices across ages

Figure 4 shows that the slot remains similar.\(^5\) The larger confidence interval for the 28–51-year-olds is partly due to the outlier.

![Figure 4](image)

**Figure 4:** The conditioning plot is drawing on the same data as Figure 3 but separated into four quartiles according to age. The age borders of separation are the 1st quantile, the median, and the 3rd quantile from project FR&DE1.

When ignoring the outlier, Pearson’s product-moment correlation coefficient remains similar across the four age groups \( (r_p; \text{between 0.85 and 0.89}) \). The oldest

---

\(^5\) We used a locally estimated scatterplot smoothing in order to explore graphically the trend in the data without forcing a linear relation. The loess method requires a large amount of data, which is clearly not the case in the older group. As a consequence, the curve is locally directed downwards due to the large numerical differences on the y-axis in the five rightmost datapoints of the graph. This result may be due to chance.
age group has the widest 95% confidence interval [0.63–0.94]. The 22–27-years-olds group has the thinnest [0.81–0.91]. They correspond to the group with the smallest and the largest participant numbers.

### 4.1.3 The relation between the LexTALE and BLP scores for each language

The trends in Figures 3 and 5 are comparable. In Figure 5, the trend line is less steep and the strength of the relation weaker for German ($r_s$: 0.62) than for French ($r_s$: 0.81).

![Figure 5](image)

*Figure 5:* The relation between the LexTALE score (from 0 to 100) and the BLP score (from 0 to 218) for each language. Each participant is represented with a point, once in (a) and once in (b).

### 4.2 Question 2

How are the components of the dominance construct – language history, use, proficiency, and attitudes – related to the vocabulary test LexTALE?

First, we check the linear association between the language proficiency score collected through the BLP and the score obtained via LexTALE graphically and with a correlation analysis. Second, we compute linear regressions with all four components of the BLP as predictors of vocabulary competence.
4.2.1 The relation between self-evaluation of language proficiency and the LexTALE scores in each language

According to Figure 6, high values in the self-assessment tend to correspond to high values in the vocabulary test. No participant got 100% correct answers in any version of LexTALE. The mean (79; 80) and the median (81; 81) of the LexTALE_FR and LexTALE_DE scores are very close as well as the mean (21; 22) and the median (23; 23) of the French and German proficiency scores as measured by the BLP.

For German, the data are more skewed due to the high number of participants assessing their language proficiency at the maximal level (112 participants in German vs. 85 in French), mainly in the project FR&DE1. Despite a positive and substantial correlation ($r_s$: 0.66 for German and 0.74 for French), there is a lot of variation in the self-assessment scores for the same LexTALE score.

![Figure 6](image.png)

**Figure 6:** The relation between self-assessment of proficiency in the BLP (from 0 to 24 points) and vocabulary proficiency in LexTALE (from 0 to 100 points) in French and German separately. Each participant is represented by two points, one in (a) and one in (b). We added horizontal random noise to the points to avoid overplotting. The black line summarizes the trend. The 95% confidence band is in grey.

4.2.2 Relation between the four BLP components and LexTALE results

To gain a better understanding of the relationship between the BLP and both LexTALE, we look at language history, language use, language proficiency, and language attitude as separate predictors for LexTALE results.
The four components of the BLP are related to each other to a certain degree (see Figures 12, 13, and 14 in the Appendix at the end of the paper). This was expected since a person who learns language A earlier and uses it more often than language B is likely to be more proficient in language A than in language B. Nevertheless, there were no dubious multicollinearity patterns between the variables for the regression models reported below (variance inflation factor – VIF – values between 1 and 4).

**Analysis of the subtraction-derived indices.** How do the different dominance components explain LexTALE indices? After analyzing several linear regression models (with or without statistical control for age or project, cf. online R-script), we report a simple linear model based on data without the outlier.

There is a positive association between each component of the BLP and the LexTALE index. Proficiency is the strongest predictor (estimated coefficient: 0.56 ± 0.1 \( \rightarrow \beta \pm SE \)), followed by language history (0.21 ± 0.4.) and language use (0.13 ± 0.3). The effect of language attitude (0.07 ± 0.06) is not significant (Figure 7).

**Figure 7:** Effect plots of each BLP component as a predictor variable for the subtraction-derived LexTALE index while holding constant all other predictor variables (Fox and Hong 2009). Lines on the x-axis represent participants. Adjusted \( R^2 \): 0.78.

**Analysis of the German and French scores separately.** The way people completed LexTALE (paper-pencil vs. online version) could have influenced the manner they filled in the survey. After fitting several models per language (cf. online R-script), we report simple linear models based on the data without the
outlier and where the variable “project” was integrated as a predictor. Here, the factor “project” affects the relationship between the subtraction-derived BLP and LexTALE indices. For the same score in the BLP components, the participants from project FR&DE1 tend to get better LexTALE_FR and LexTALE_DE results than the participants in project FR&DE2. This is counterintuitive since participants from project FR&DE2 theoretically could correct their answers. Since the variable “project” is less important for the answer to the research question, we limit the following discussion to the association of the four components of the BLP with the LexTALE score.

Figure 8: Effect plots of each BLP component as a predictor variable for the LexTALE German scores while holding constant all other predictor variables, i.e. the other BLP components and the variable project. Lines on the x-axis represent participants. Adjusted $R^2$: 0.55.

6 Since the factor project only has two levels, it was modelled as a fixed effect and not a random effect. This can limit the generalization of the results above these two projects (Yarkoni 2019), but the big picture regarding the order of importance for the four BLP components as a predictor for LexTALe scores remains the same across the models with or without project as a variable (cf. online R-script).
Again, proficiency is the strongest predictor (estimated coefficient for German: $0.57 \pm 0.1$; for French: $0.43 \pm 0.1$). For French, language use is the second most important predictor ($0.28 \pm 0.05$) whereas it is not significant for German ($0.06 \pm 0.04$). Language history is significant for German ($0.19 \pm 0.05$) and French ($0.2 \pm 0.05$). Language attitude remains the weakest (and non-significant) predictor (Figures 8 and 9).

4.3 Question 3

Is the relationship between the BLP and the lexical decision task in French/German bilinguals mirrored in data from Italian/German bilinguals completing the Dialang test instead of LexTALE for Italian?

The BLP as a measure of language dominance, LexTALE as a measure of German vocabulary proficiency and VSPT (Dialang) as a measure of Italian
vocabulary proficiency were used in the project IT&DE. Both lexical decision tasks’ maximum score is 100. Figure 10 shows the distribution of the BLP indices and proficiency indices. The latter was obtained by subtracting the Dialang scores in Italian from the LexTALE scores in German.

The density curve in Figure 10a has a comparable shape to that in Figure 2a with the French/German bilinguals. Data in Figure 10b are unequally distributed around zero because participants tend to get higher scores in the VSPT from Dialang_IT than in LexTALE_DE. If we look at the scores separately, the mean and median scores for the Italian test (∼90/100) are about 10 points higher than for German and French (∼81/100). There is barely any ceiling effect since no participant obtained 100 points, but data are shifted to the left for Italian.

![Figure 10](image_url)

**Figure 10:** Density plot of BLP (a) and LexTALE_DE/VSPT from Dialang_IT subtraction-derived indices (b). The x-axis of (a) represents the dominance continuum: the further to the left, the more participants are considered dominant in Italian; the further to the right, the more they are considered dominant in German. (b) Shows the repartition of the participants according to their proficiency: the further to the left, the better their proficiency score in the Italian test in comparison to the German test and vice-versa.

Figure 11 shows that the slope goes in the same direction but is less steep than in Figure 3. In contrast to Figure 3, the coordinate \( y = 0 \) and \( x = 0 \) lies outside the confidence band.
The correlation between the two subtraction-derived indices is lower \((r_{p,s}: 0.63)\) than for the French/German bilingual data \((r_s: 0.86)\). To compare the predictability of the BLP index according to the vocabulary proficiency index, we ran a simple linear regression:

\[
\text{Predicted BLP index} = 36.6 + (3 \pm 0.5) \times \text{vocabulary index}.
\]

Standard deviation = 45.9. Multiple \(R^2 = 39\%\).

The observations based on the comparison between Figures 3 and 11 are confirmed by the regression analysis. The model for the IT&DE data has a weaker predictive/explanatory power (see \(R^2\)) and has more uncertainty than the two models based on data from French/German bilinguals. The slopes of the vocabulary index in IT&DE is a few tenths lower than that for French/German bilinguals.

5 Summary and discussion of the results

In what follows, we synthesise the results of each research question and discuss possible explanations or interpretations.

**Question 1**: The data on French/German bilinguals showed a strong linear relationship between vocabulary proficiency (LexTALE) and language dominance.
(BLP) indices. LexTALE_FR and LexTALE_DE produced a fairly comparable result in our sample if we take the BLP as a benchmark.

The relationship between the two measures is stronger for French than for German. A possible explanation is that the questions about (Swiss) German in the BLP are less valid or discriminate less well for some participants that use dialect by default in comparison to language users who mainly use a standard variety.

As suggested by an anonymous reviewer, the LexTALE_FR could have been easier than the LexTALE_DE, since the French version also targets those with little knowledge of French. The mean and the median scores of the LexTALE_FR and LexTALE_DE are very close in our data. Given the analysis in part 3.3, 4.1, and 4.2.1, we cannot claim that the LexTALE_FR is easier than LexTALE_DE, at least for our participants who are fluent in both languages.

**Question 2:** First, we compared the scores from the self-assessment of proficiency from the BLP with the LexTALE scores in each language separately. The correlation coefficients (around 0.7) were comparable to the relationship between LexTALE and the whole BLP for each language, showing a good degree of association. We observed a ceiling effect for the self-assessment measure, especially for German. The absence of a ceiling effect for LexTALE_DE and LexTALE_FR suggests that lexical decision tasks discriminate in a more fine-grained way among highly proficient speakers than self-assessment. Second, we analyzed the relation between the LexTALE results and the four weighted components of the BLP. Regression models were fitted for the subtraction-derived indices as well as for the scores in each language separately. In all models, the proficiency component explains, as expected (see point 2.2), most of the variance in LexTALE results. In contrast, the attitude component never yields a significant estimate. Maybe this component corresponds to a known unknown, the “something other” than proficiency skills that is targeted by specific operationalizations of language dominance.

Language use was a significant predictor for the subtraction-derived indices as well as for the French scores, but not for German scores. The fact that participants were asked to consider the dialectal and the standard varieties together could have influenced this result. Another explanation is that language use can be unstable. Most of the data were collected in the bilingual town of Fribourg, in which French is the majority language (Coray and Berthele 2018). Thus, participants who are more proficient in German may have a more balanced use of the languages than participants who are more proficient in French. Moreover, when looking at the raw score for language use, we observe that some participants pretend almost not to
use one of their languages, even if they all could fluently participate in a retelling task (see the numerical summary of the data in the online script).⁷

Posthoc simple regression analysis showed the coefficient for the language use in French (0.16 ± 0.1) explained more variance in the self-evaluation of proficiency in French ($R^2 = 48\%$) than the coefficient for the language use in German (0.13 ± 0.1) explained in the self-evaluation of proficiency in German ($R^2 = 40\%$). However, taking into consideration measurement errors and the small size of the difference, this result says little about the relation between language use and language proficiency measures in each language.

**Question 3:** To investigate the generalizability of the findings from question 1 to other language pairs or test combinations, we draw on a data set of Italian/German bilinguals. An analysis of the subtraction-derived indices showed a positive relationship between the BLP and the lexical proficiency measures. As expected (see point 2.2), this relation is positive and linear but also weaker than for the French/German bilinguals. However, this may be partly due to the fewer number of participants in the Italian/German bilinguals sample in comparison to the French/German bilinguals sample. Balanced bilingualism according to the BLP differed from a balanced proficiency according to the two lexical decision tasks. On average, participants got higher scores in the test in the VSPT form Dialang_IT than in LexTALE_DE.

### 6 Conclusion and implications

This study addressed methodological and conceptual problems frequently encountered in quantitative studies investigating bilingualism: What measures are used to shed light on bilinguals’ language dominance and how comparable are these different ways of operationalizing the bilinguals’ language profiles? We now discuss how our findings can contribute to dominance-related research in bi- and multilingualism.

From a theoretical point of view, this result suggests that dominance indices based on a multi-components survey and indices based on receptive vocabulary tests may measure a similar construct. LexTALE scores not only represent a measure of the receptive vocabulary of a speaker but are also a proxy for general language proficiency (cf. Lemhöfer and Broersma 2012). The patterns in our data make sense since the language skills in both languages of a bilingual should be

---

⁷ A reviewer of this paper raised the question whether the language use of the BLP has been too generic. Since we did not compare the language use module of the BLP with a more fine-grained measure of language use in different domains, we cannot answer this question.
related to the weight of these languages in his or her life, at least in terms of history and usage. As discussed under point 5, the linear relationship between the LexTALE scores and the language attitudes component of the BLP is very weak. This suggests that language attitudes may indeed be one of the components that differentiate language dominance defined in holistic terms from skill-focused language dominance. However, our study can only rise this hypothesis as we didn’t take measurement errors as well as other measures of language attitudes into account. This brings us back to the question of construct validation.

It remains unclear what results should be expected to validate an instrument. Lemhöfer and Broersma (2012) denounce the lack of a recognized standard vocabulary test that could be used to validate a new instrument like LexTALE. This problem applies to language dominance as well.

Methodologically, our results drawing on BLP and LexTALE scores and indices provide useful information. When time is short for data collection, one could choose one of these two measures to operationalize language dominance.

The BLP has the advantage of being usable with more language pairs. Researchers often create their survey instruments. One frequent argument is that the way to operationalize dominance depends on their immediate research purposes (cf. review in Treffers-Daller 2015: 240). However, including a widely shared, usable, and freely available instrument in studies tapping into language dominance not only facilitates methodological decisions in the planning stage of the research; it also produces results that are more readily compared across contexts and languages. Besides, researchers can/should also differentially weight the BLP items, use the individual components separately or use the raw score in each language to suit their specific purposes (cf. Birdsong 2015).

Despite arguments against self-evaluation of the four language skills (cf. Treffers-Daller 2015: 254), the proficiency module of the BLP can be considered as a good proxy when looking for a subtraction-derived index for language/lexical proficiency (see also Figure 12). When interested in language proficiency in each language separately, LexTALE_FR and LexTALE_DE provide more precise results.

Replication studies with different language combinations and language proficiency tests as well as with participants with less balanced bilingualism will be necessary to assess the generalizability of our insights.

**Acknowledgments:** All projects are financed by the Swiss National Science Foundation (SNFS) and granted to Raphael Berthele: Information structure in bilingual speakers: Cross-linguistic influences and language dominance (2017–2021) (project number-176338) and Motion at the linguistic border. On the influence of language dominance and language mode on the expression of motion in
bilinguals (*French-German*), (project-number: 156121). A depth of gratitude is owed to Giulia Berchio who allowed us to analyze her data (project IT&DE). Many thanks also to Jan Vanhove for discussing the analysis in an earlier step of this study.

**Competing interests:** The authors declare that they have no competing interests.

### Appendix

**Figure 12:** Pair plot of the relationships between the subtraction-derived LexTALE index and each BLP component index for the French/German bilingual data.

**Figure 13:** Pair plot of the relationships between the LexTALE_DE score and each BLP component score on German for the French/German bilingual data.
References


