

Berthele, R., & Vanhove, J. (2020). What would disprove interdependence? Lessons learned from a study on biliteracy in Portuguese heritage language speakers in Switzerland. *International Journal of Bilingual Education and Bilingualism*, 23(5), 550–566. <https://doi.org/10.1080/13670050.2017.1385590>

KEYWORDS Interdependence hypothesis; thresholds; biliteracy development; crosslinguistic influence; transfer

Abstract

This article reports on an investigation into the bilingual development of literacy skills in Portuguese heritage language speakers in Switzerland in which we sought to put Cummins' linguistic interdependence hypothesis to the test. Three predictions were derived from this hypothesis: (1) Literacy skill levels in one of the bilinguals' languages can best be predicted from skill levels in both languages at an earlier point in time; (2) Crosslinguistic effects originating from the language in which literacy skills are explicitly taught are stronger than those originating from the language in which they are not explicitly taught; (3) Crosslinguistic effects are stronger between more closely related language pairs. These three predictions were evaluated using longitudinal reading and writing data from 233 French–Portuguese and German–Portuguese bilingual children. Results show that prediction (1) is borne out but predictions (2) and (3) are not. The discussion centres on the vast array of possible results that are compatible with the interdependence hypothesis in correlational studies, particularly if it is combined with *post hoc* thresholds. Our conclusion is that current forms of the interdependence hypothesis do not offer a fruitful framework for further investigation.

Background

This article reports on a longitudinal study that investigated the bilingual development of literacy skills in bilingual heritage language speakers in Switzerland. At the outset, this study aimed to assess whether and how strongly these speakers' languages are interdependent. With the benefit of hindsight, however, it is unclear what kind of results would indicate that bilinguals' language are *not* interdependent, as we will further discuss in this introductory section. Nonetheless, we think that the study's findings and the conceptual and methodological lessons we have drawn from it are worth reporting. In our concluding discussion, we critically assess the potential of interdependence-related thinking for the fruitful investigation of bi-literacy development.

Cross-linguistic influence and interdependence

The idea that bi- and multilinguals' languages influence each other is one of the tenets of contemporary scholarly work in our field. On the level of linguistic features, researchers in the wake of Weinreich's (1953) seminal work have contributed to a better knowledge of the directions and constraints

CONTACT Raphael Berthele raphael.berthele@unifr.ch

on cross-linguistic influence and of how this type of transfer can be investigated empirically (Jarvis 2000; Jarvis and Crossley 2012; Jarvis and Pavlenko 2008; Odlin 1989). Apart from this linguistic take on multilingual repertoires, educational scientists have argued that another type of transfer, in particular in relation to literacy acquisition in bi- and multilingual children, needs to be considered. This type of transfer is often referred to with the term 'linguistic interdependence', whose definition by Cummins is as follows:

To the extent that instruction in Lx is effective in promoting proficiency in Lx, transfer of this proficiency to Ly will occur provided there is adequate exposure to Ly (either in school or environment) and adequate motivation to learn Ly. (Cummins 1996, 111)

Transfer of proficiency thus depends on a) efficiency of Lx instruction, b) exposure to Ly and c) motivation to learn Ly. This transfer across languages, according to Cummins (2009, 2017), involves conceptual elements (i.e. world knowledge), linguistic elements (e.g. cognate words), metalinguistic awareness (involving phonology, morphology, syntax), pragmatics as well as metacognitive elements (e.g. learning strategies). We do not doubt that some of these elements indeed are 'transferred', for example we do not doubt that understanding photosynthesis and having words for it in Lx can be 'transferred', that is talked and thought about in Ly, provided the linguistic means for doing so are acquired. This type of 'transfer' (if transfer is the right word), in our view, is uncontroversial and does not call for empirical investigation. Some scholars argue that 'transfer or the existence of underlying attributes based on cognitive and personality attributes of the individual' are 'two sides of the same coin' (Cummins 2017, 107). The main problem in extending the notion of transfer to cognitive attributes in general is that in that case the theory should no longer be considered one of *linguistic* interdependence, but of general cognition. Therefore, correlational patterns of measures of linguistic skills would be mere epiphenomena of general cognitive effects and the educational discussion should no longer be about language promotion but about promotion of cognitive skills. It seems to us that such a wide construal of transfer and interdependence undermines the whole point of the interdependence-based argument in favour of an instrumental value of bilingual education. The question that we address is thus a more specific form of transfer, namely to what extent multi-lingual learners can benefit from literacy related skills acquired in one language in their other language(s).

Cummins' work on interdependence and the common underlying proficiency (CUP) is to be understood as a response to alternative explanations that view language learning from a time-on-task perspective. On these explanations, a first language (L1) other than the school language represents an impediment to literacy development and educational success (see Cummins 1996, 112, and Baker (2006), for more details on the history of biliteracy research). Cummins' work had and still has a considerable impact on policies in many different contexts, in particular with respect to bilingual education in (immigrated) minority languages and the local school languages, for example in the policy documents issued by the Council of Europe (Beacco et al. 2010, 57). These documents in turn influence national and other supranational policies.

The interdependence view on bi- and multilingual literacy acquisition has now become common currency in the field to such an extent that only few question its role in scholarly thinking about multilingual competence. However, as we will argue in this contribution, there are a number of theoretical and methodological questions remaining, and depending on how these questions are resolved, the notion of interdependence is either unfalsifiable or in its wide application to all kinds of shared or transferred cognitive attributes neither a useful nor a convincing argument in the context of bilingual education.

Investigating the common underlying proficiency

Faced with conflicting claims about linguistic transfer on the level of linguistic features, researchers (most notably Jarvis 2000, 2010, 2012) drew up a framework for evaluating empirical evidence in

favour of or against such transfer. Such a framework is currently lacking for the investigation of the more hidden working of the CUP. The CUP corresponds to Shuy's (1978) 'hidden part of the iceberg' and involves important components of Cummins' CALP (cognitive academic language proficiency) dimension of language. Hulstijn's (2015) distinction between Basic and Higher Language Cognition (BLC-HLC) is not identical but nonetheless related to this, since HLC involves higher-level cognitive operations (e.g. 'reflective, meta-linguistic awareness', Hulstijn 2015, 61). Transfer on this level has been investigated in terms of (reading or writing) strategies, metacognitive knowledge (about reading or writing) or more general reading/writing ability (see for instance García, Jiménez, and Pearson 1998; Idiazabal and Larrangan 1997; Muñiz-Swicegood 1994; Schoonen, Hulstijn, and Bossers 1998; van Gelderen et al. 2007), but certain problems remain with how data garnered in such empirical investigations is used to assess linguistic interdependence.

Group comparisons

One common design for investigating the CUP is to compare the linguistic development of children receiving bilingual instruction or first language support to that of children not receiving such instruction or support (see our discussion section and, e.g. Reljić, Ferring, and Martin (2015) for a meta-analysis). The children in such studies are, for ethical and practical reasons, rarely randomly assigned to treatment and control groups. Instead, they, or their parents, choose which programme they wanted to attend, often based on some formal or informal assessment of the child's general and language learning skills. Inferences from such studies on the whole population of bilingual children should be made with caution. In the US context, Rolstad, Mahoney, and Glass (2005b) discuss a number of earlier reviews and meta-analyses that included only randomised programme evaluations and which paint a mixed picture of the programmes' effectiveness in fostering L2 proficiency through L1 support. In their own meta-analysis, which did include non-randomised evaluations, Rolstad, Mahoney, and Glass (2005b) found that long-term bilingual programmes for English-language learners in the US yield better proficiency in both English and the first language than do English-only programmes (see also Rolstad, Mahoney, and Glass 2005a); the authors did not speculate about any transfer/interdependence mechanisms behind this effect.

Covariation patterns at a specific point in time

Another common research design for investigating the CUP is a correlational one in which the researchers assess the association of skill measured in one language with the same or similar skills in the other language at the same point in time in a sample of bilinguals (e.g. Da Fontoura and Siegel 1995; Proctor et al. 2010). In principle, such studies could falsify the interdependence hypothesis if they do not show any strong association between the bilinguals' skills in the two languages. However, if no such associations are found, for instance, if L1 reading does not correlate with L2 reading, scholars can take recourse to threshold-based explanations. An example for such an interpretation of low correlation coefficients as proof of thresholds in reading is Lee and Schallert (1997; see Takakuwa 2005 for more examples and a critical discussion of the threshold theory) (Of course, this is not to say that all scholars appeal to some threshold to explain low correlations since one can hold a view of interdependence without thresholds.). According to threshold explanations, transfer and interdependence are constrained by the participant's language proficiency in the sense that they need to reach a certain proficiency threshold if they are to benefit from the advantages of additive bilingualism. This view held sway in societal debates and led to linguistic policies that aimed to foster one or the other language. Of note, various threshold hypotheses exist, and these differ with regard to the language in which the child needs to reach a certain proficiency level in order to benefit from positive transfer (L2 for Alderson 1984; and Kecskes and Papp 2003; both L1 and L2 for Cummins 1979). Moreover, they differ with regard to the requisite kind of linguistic knowledge (in particular: surface vs. underlying elements).

The difficulty with threshold-based explanations is to define these thresholds theoretically and to provide empirical evidence for their existence (for a discussion, see Hall, Cheng, and Carlson 2006;

Takakuwa 2005). As Cummins (1979) notes, '[t]he threshold cannot be defined in absolute terms; rather it is likely to vary according to the children's stage of cognitive development and the academic demands of different stages of schooling' (230). While it may well be true that thresholds exist and that they are both relative (inter-individual) and dynamic (intra-individual), this would do little to inform policymaking since a reform well-suited for one child would not necessarily have positive effects for another child (see for instance Takakuwa 2005).

To date, the thresholds are rarely defined or operationalised a priori. What is more, no one so far has shown empirically that there is a non-linear developmental slope whose shape would be evidence for such thresholds (for a related discussion, see Vanhove 2013, Vanhove 2014).¹

Longitudinal covariation patterns

To the extent that group comparisons and covariation patterns at a specific point in time can inform theorising about interdependence and cross-linguistic transfer, they often do not permit any inference about the causal relationships between the linguistic skills in the different languages and other cognitive skills and academic achievement. In order to better establish causal patterns (instead of purely correlational patterns), several studies on bi-literacy have therefore investigated transfer using longitudinal designs. In these studies, researchers investigate so-called cross-lagged correlations, that is, they assess how useful the performance of an individual at time T in one language is for predicting their performance at time $T + 1$ in the other language (e.g. Cárdenas-Hagan, Carlson, and Pollard-Durodola 2007; Verhoeven 1994). A crucial but as yet unresolved question is how long the time interval between the two measurements should be. Ideally, the choice of the time lag is based on theoretical considerations and is taken into account when interpreting the study's results (Collins 2006; Collins and Graham 2002; Selig and Little 2012). To our knowledge, however, there is no theoretical or empirical basis for determining the time interval it takes for a particular subskill developed and matured in one language to be carried over to the other, and establishing such a basis is a clear desideratum for future theoretical work. As we will point out in the discussion of our own study below, however, even theoretically arbitrary time-lags need to be taken into consideration when interpreting the results.

Interim discussion: explanations for covariation patterns

Having outlined the main research designs for investigating interdependence, we will now discuss what results such designs could produce and how such results could be interpreted. If we had the luxury of unlimited and high-quality data, the relationship between skills in two languages, the results of a correlational study (be it longitudinal or not) would have to correspond to one of these four general patterns:

- a) a negative linear association of L1 and L2 skills;
- b) no association between L1 and L2 skills;
- c) a positive linear association of L1 and L2 skills;
- d) non-linear associations (modifying A or C, e.g. due to thresholds)

Let us consider, for each of these patterns, whether it is consistent with the view that L1 and L2 are interdependent and whether alternative explanations can be found for it.

As for patterns (A) and (B), such nil or negative correlations would fit in with what Cummins, in his earlier work, referred to as 'semilingualism' and to what is sometimes called 'subtractive bilingualism' (Cummins 1976). While these patterns are incompatible with a threshold-free versions of the interdependence hypothesis, in terms of a thresholded interdependence hypothesis, they could signal that the bilinguals below the proficiency threshold required for positive cross-linguistic effects to kick in. Patterns (A) and (B), however, are also compatible with the *time on task* idea, that is, the idea that the growth of the linguistic system is mainly a function of the time spent on task in that language and that the inter-lingual benefits are negligible (Esser 2009).

As for pattern (C), positive correlations are compatible with both thresholded and threshold-free versions of the interdependence view; in thresholded versions, such correlations would signal that the bilinguals have reached the threshold required to profit from positive cross-linguistic effects. But several other explanations would be compatible with this pattern, too. Such explanations include language giftedness (see Bishop et al. 2006; Rimfeld, Dale, and Plomin 2015; Stromswold 2001 on the genetic dimension of language giftedness) and reliance on general cognitive skills such as working memory, general intelligence, and world knowledge.

Finally, pattern (D) could be compatible with a thresholded interdependence view, but some non-linear relationships can be caused by floor and ceiling effects as well. In sum, positive associations of language skills can be explained by several mechanisms that are not necessarily mutually exclusive (interdependence, language giftedness, general cognitive abilities). But whereas a general interdependence theory without thresholds is incompatible with no or a negative association, no imaginable data pattern is incompatible with the interdependence theory plus *post hoc* thresholds. In the remainder of this article, we formulate three predictions we derived from the interdependence framework and discuss how we put these predictions to the test.

Testing interdependence-based predictions using data from biliterate children

On the HELASCOT project

The *Heritage language and school language: are literacy skills transferable?* (HELASCOT) project has been conceived as part of the 2011–2015 research programme of the Research Centre on Multilingualism (Institute of Multilingualism, Fribourg). The main goal of the HELASCOT project was to put the interdependence hypothesis to the test: Is there empirical evidence for the cross-linguistic influence of one language of heritage language speakers on the development of the other language(s) in their repertoire. More precisely, the HELASCOT project aimed to describe the development of literacy skills in Portuguese heritage speakers in Switzerland from the beginning of Grade 3 (roughly 8 years of age) to the end of Grade 4 (10 years of age) in both of their languages, and to document instances of the cross-linguistic transfer of literacy skills. In the analyses presented in this article, we will focus on reading comprehension. However, the study also investigated the development of writing proficiency. The writing data yielded a similar picture to the reading data overall; where there were differences, this will be noted below.

Three predictions derived from the interdependence hypothesis

We assessed the following three predictions.

1. An individual's score at time T predicts their score in the same skill in the other language at time $T + 1$.

If languages are interdependent, we expect to find that good (weak) readers in the heritage language (HL) at one point in time will be relatively good (weak) readers in the school language (SL) at a later point in time. Similarly, good readers in the SL are expected to be relatively good readers in the HL at a later point in time. Crucially, we will already be able to predict with some accuracy the children's performance in a given language at time $T + 1$ from their performance in the same language at time T , and such within-language effects need to be taken into account, too. The crucial question, then, is: 'To what extent does an individual's score at time T predict their score in the same skill in the other language at time $T + 1$ after taking into account their score in the same skill in the other language at time T ?'. The prediction is that positive cross-linguistic longitudinal effects exist even after taking into account within-language effects.

2. Crosslinguistic effects from the school language to the heritage language are stronger than the other way around.

The HELASCOT participants received most of their literacy education in the SL – even though most participants also took HL courses (see Lambelet et al. 2017). This prediction, as far as we know, is nowhere explicitly stated in the interdependence literature. However, based on Cummins' definition of interdependence as cited above, we expect transfer to be stronger in the direction from SL to HL, since the children get sustained literacy instruction at school, whereas the literacy practices outside of school, even if a couple of hours a week are dedicated to heritage language instruction, are rather limited. The interdependence-related expectation would therefore be that literacy skills developed in the SL will be put to use in the HL, in which less literacy training is received. We therefore expect crosslinguistic longitudinal effects to be stronger from the SL to the HL than from the HL to the SL. This expectation is not at odds with the first prediction: while we expected SL-to-HL effects to be stronger, this does not imply that we expected that there would be no HL-to-SL effects. Furthermore, if this prediction were to be confirmed, it would be more difficult to account for it in terms of the alternative explanations we discuss above.

3. Crosslinguistic effects between French and Portuguese are stronger than those between German and Portuguese.

Two SLs are involved in the HELASCOT project: French and German. French is a Romance language that is genealogically closely related to the participants' HL (Portuguese). German, on the other hand, is a Germanic language and more distantly related to the HL. Sometimes, literature inspired by the idea of interdependence seems to show that interdependence is not or only weakly dependent on linguistic proximity of the languages involved (e.g. Dressler and Kamil 2006). Other scholars investigating interdependence, however, acknowledge the influence of linguistic proximity on interdependence effects (Proctor et al. 2010). Furthermore, studies on Second Language Acquisition (e.g. Schepens 2015), but also from transfer research in non-linguistic domains (Singley and Anderson 1988), strongly emphasise that the number of overlapping elements (viz. the degree of linguistic relatedness) is a robust predictor of learning gains.

Thus, the greater formal (lexical and morphosyntactic) similarities between French and Portuguese are expected to be more conducive to positive cross-linguistic effects, and hence stronger crosslinguistic longitudinal effects are expected between French and Portuguese than between German and Portuguese. Like Prediction 2, Prediction 3 is not at odds with Prediction 1, and confirmatory evidence for Prediction 3 would not be as easy to account for in terms of the alternative explanations discussed earlier.

Method

Sample

The participants were children with Portuguese as a heritage language living in Switzerland. 114 of these children lived in the French-speaking part of Switzerland and had French as their school language; 119 lived in the German-speaking part of Switzerland and had Standard German as their school language. 73% of the Portuguese participants in French-speaking Switzerland and 97% in German-speaking Switzerland took part in heritage language classes with a focus on literacy development (generally 2 h per week). Parents reported to engage in literacy practices (e.g. reading) between once a week and every day, the average number of books in the Portuguese families is low (between 10 and 20 books; cf. Desgrippes and Lambelet 2017 for details on the home literacy environment). Our original idea was to compare children with and without heritage language instruction longitudinally, but this plan had to be abandoned given the low proportion of Portuguese

children in Switzerland who do not take part in such classes. 85% of the children in the sample were born in families that had already immigrated before their compulsory school started. Additionally, three groups of children without Portuguese as a heritage language served as comparison groups: 78 in French-speaking Switzerland, 80 in German-speaking Switzerland and 91 in Portugal.

These children were tested at three points in time. The first data collection took place at the beginning of Grade 3 (average age: 8 years and 8 months), the second at the end of Grade 3, and the third at the end of Grade 4 (average age: roughly 10 years). Due to subject unavailability, not all children were tested at each data collection; the figures in the Results section mention the number of available children per data collection.

For more details about the recruitment of the participants, see Lambelet et al. (2017). For a full-fledged description of the participant sample, we refer to Desgrippes and Lambelet (2017).

Task

The HELASCOT project investigated the development of both the children's writing and reading skills. The writing and reading data yielded highly similar results as far as the three predictions outlined above are concerned. In the interest of space, we therefore only discuss the reading data in this contribution. Readers interested in the writing tasks are referred to Lambelet, Desgrippes, and Vanhove (2017) and Vanhove and Berthele (2017).

The reading tasks, whose construction, administration and scoring are described in more detail in Pestana, Lambelet, and Vanhove (2017), consisted of a translated extract from Lewis Carroll's *Alice's Adventures in Wonderland* and a series of 13 related multiple-choice questions and two open questions related to explicit information, implicit information, global coherence, and emphatic reading. In the Portuguese version, Alice has just fallen down the rabbit hole and starts pursuing the rabbit. In the French and German, the rabbit takes Alice for a housemaid and sends her into his house to fetch him gloves and a fan. The Portuguese-French and Portuguese-German bilinguals completed both the Portuguese and the French or German reading tasks; the children in the comparison groups only completed the reading task in their respective language. The same reading tasks were used at *T* 1, *T* 2 and *T* 3. This allows for a direct comparison of the scores without having to scale different tests and items on an independent proficiency scale. Across all three data collections, all languages, and all participant groups, the reading tasks' 21 reliability coefficients (Cronbach's α) varied between 0.65 and 0.84 and averaged 0.74 (see Pestana, Lambelet, and Vanhove 2017). The global scores on which the following analyses were based are the percentages of correct responses across all items.

Results

Before testing the three interdependence-based predictions, we briefly describe the distribution of the reading task scores across the data collections, groups, and languages as well as the intercorrelation between scores at one point in time and those from a year later.

Reading skills in the school and heritage language

Figure 1 shows box plots for the reading task scores, converted to percentages. The boxplots show that the bilinguals' scores in French and German tend to be lower at *T* 2 and *T* 3 compared to the comparison groups (visible in the lower medians and central box containing the middle 50% of the data points). Note that the French, German and Portuguese reading tests were not normed on the same scale. While the French, German and Portuguese tests were structurally identical, and the French and German tests were translations of each other, this does not imply that the same result on the different tests attest to the same skill level. For this reason, it is inadvisable to directly compare, say, the scores in French to those in German, so the relatively low scores in German by bilinguals and comparison

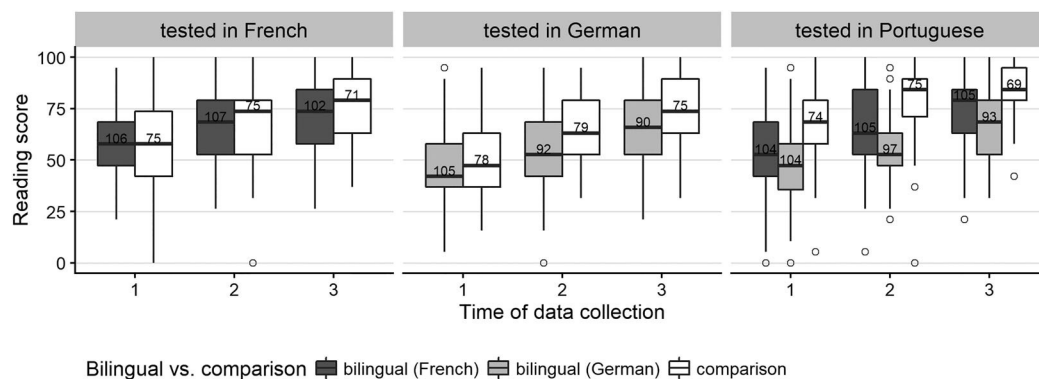


Figure 1. Reading comprehension scores of participants with Portuguese as a heritage language in French- (dark grey) and German-speaking Switzerland (light grey) and of natively French-, German- and Portuguese-speaking children (white) at data collections 1, 2, and 3. The numbers of participants are given in the boxplots.

participants alike should not be taken to suggest that children based in German-speaking Switzerland had lower literacy levels than those in French-speaking Switzerland.

Within- and between-language longitudinal correlations

Figure 2 shows how the bilinguals' French, German, and Portuguese reading skills at T_2 and T_3 correlate with their French or German reading skills at the previous data collection (T_1 and T_2 , respectively). As expected, the French–French and German–German correlations are positive: relatively good readers in French/German tend to be relatively good readers in French/German one year later. The between-language correlations (French–Portuguese and German–Portuguese) are also positive, and a striking observation is that the weak relationship between German and Portuguese between T_1 and T_2 (upper right panel) seems to be considerably stronger between T_2 and T_3 (bottom right panel).

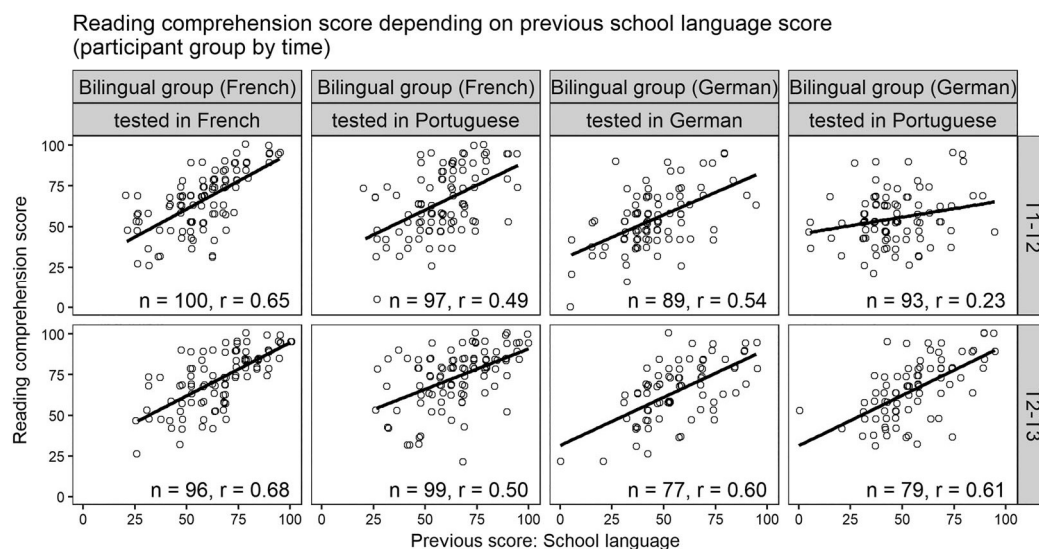


Figure 2. Relationship between reading comprehension scores in the school language at one data collection and reading comprehension scores in French, German, and Portuguese at the next data collection for participants with Portuguese as a heritage language. Upper row: school language at T_1 and both languages at T_2 ; bottom row: school language at T_2 and both languages at T_3 .

Figure 3 shows how the bilinguals' French, German, and Portuguese reading skills at $T2$ and $T3$ correlate with their Portuguese reading skills a year before ($T1$ and $T2$, respectively). As expected, the Portuguese–Portuguese correlations are positive: relatively good readers in Portuguese tend to be relatively good readers in Portuguese one year later. In addition, the between-language correlations are also positive: relatively good readers in Portuguese tend to be relatively good readers in French or German one year later. These correlations are weaker for crosslinguistic effects: Both French and German reading skills are less well predicted by Portuguese reading skills at $T1$. These across-language relationships going from the HL to the SLs are stronger in the second time interval tested. As a reviewer pointed out, this latter pattern is 'noteworthy in light of considerable data (reviewed in Cummins 2001) showing that as students gain more opportunity to develop literacy in both their languages, cross-lingual relationships increase in strength.' Note, however, that the French–Portuguese correlation at $T1 - T2$ ($r = 0.49$) is essentially identical to the one at $T2 - T3$ ($r = 0.50$). Furthermore, this pattern is not consistently replicated in the writing data (reported in Vanhove and Berthele 2017), nor is it germane to our *a priori* predictions.

Testing the three predictions

Prediction 1: longitudinal crosslinguistic effects

The first prediction involves estimating participants' reading scores in a particular language (SL or HL) at a given point in time using their reading scores in the other language from the year before. Clearly, we can more accurately estimate their reading scores in a given language when we know their reading scores in the same language from the year before than when we don't; the question addressed here is whether additionally knowing their previous score in the other language permits more accurate estimates still. If the answer is *yes*, this wouldn't necessarily vindicate the interdependence hypothesis, since alternative explanations are possible (see the Discussion section). If, however, the answer is *no*, there would be no evidence for the interdependence hypothesis, unless we were to stipulate that all study participants were below a threshold that we now determine *ex post facto*.

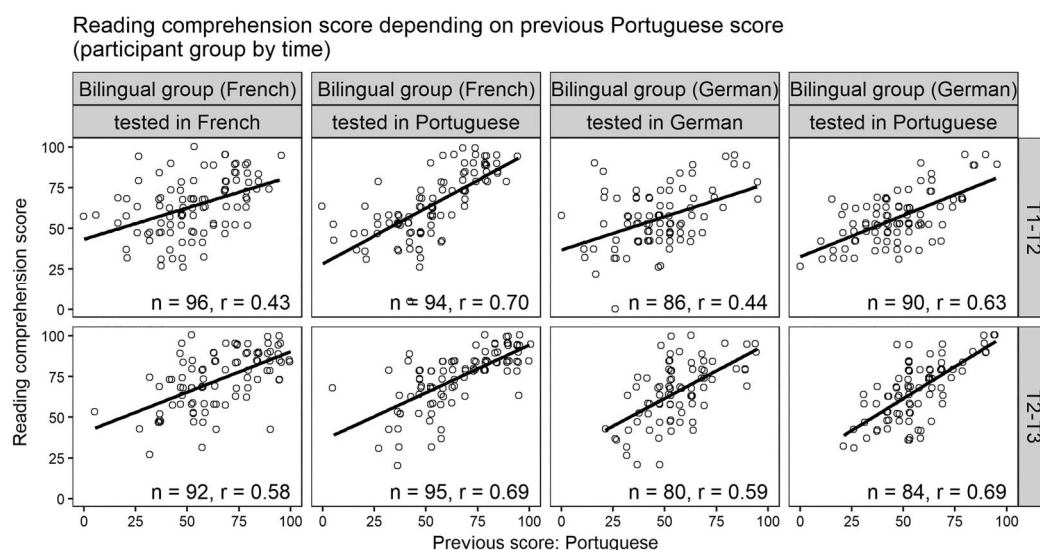


Figure 3. Relationship between reading comprehension scores in Portuguese at one data collection and reading comprehension scores in French, German, and Portuguese at the next data collection for participants with Portuguese as a heritage language. Upper row: Portuguese at $T1$ and both languages at $T2$; bottom row: Portuguese at $T2$ and both languages at $T3$.

To address this question, we analysed the data in mixed-effects models using the lme4 package (Bates et al. 2015) for R (R Core Team 2016). We briefly describe the structure of these models here, but readers interested in the fine details are referred to the online datasets and R code (<http://dx.doi.org/10.6084/m9.figshare.4861406>), whereas readers less interested in the technicalities may want to skip to the next paragraph. Mixed-effects models model an outcome variable in terms of fixed-effect predictors and random effects. In this case, the outcome variable consisted of the bilingual participants' reading scores at T_2 and T_3 in both of their languages. The fixed-effect predictors included several variables that specified the study's design. These are not of primary interest here, but we deemed it necessary to include these variables to reflect the study's design in the analyses. These fixed-effect predictors were (a) the time of data collection (T_2 vs. T_3); (b) the language tested (school language vs. heritage language); (c) language region (French- vs. German-speaking Switzerland); as well as (d) an interaction term between the language tested and language region in order to allow for the possibility that French-Portuguese bilinguals attained different scores from German-Portuguese bilinguals in either of their languages. Additionally, a fixed-effect predictor modelled the linear effect of the participants' previous score in the same language ('previous same'): if the outcome variable contained a T_2 observation in French, the corresponding predictor value was the participant's T_1 score in French; if the outcome variable contained a T_3 observation in Portuguese, the corresponding predictor value was the participant's T_2 score in Portuguese. To these fixed effects, we added a predictor that modelled the linear effect of the participants' previous score in the other language ('previous other'): if the outcome variable contained a T_2 observation in French, the corresponding predictor value was the participant's T_1 score in Portuguese; if the outcome variable contained a T_3 observation in Portuguese and the participant resided in German-speaking Switzerland, the corresponding predictor value was the participant's T_2 score in German. This is the predictor that is of chief interest. Lastly, the models contained random effects that allowed scores to vary within participants (by-participant random intercept) and within classes (by-class random intercept) as well as random effects that allowed the effects of the time of data collection, of the language tested, and of the participants' previous scores in the same and the other language to differ from class to class (by-class random slopes). These random effects were included to account for the clustered nature of the data (multiple observations per participant; participants nested in classes).

The question is whether a model with the participants' previous scores in both languages yields more accurate estimates than one with the participants' previous scores in the same language only – or in other words, whether the fixed-effect predictor for 'previous other' significantly improves the model fit. A likelihood-ratio test suggests this is the case ($\chi^2_{Lk1} = 10.7, p = .001$), that is, that there is indeed an added value in modelling the participants' previous scores in both language when estimating their scores in a particular language a year later. The estimated coefficient for 'previous other' was 0.16 ± 0.04 (\hat{b} S.E.), that is, other things equal, a participant outperforming another participant by 10 points in the other language at an earlier point in time will on average score 1.6 ± 0.4 points better in the target language at the current point in time. This compares to an estimated coefficient of 0.47 ± 0.04 for 'previous same', indicating that 'previous same' is nonetheless the stronger predictor. The full model details are available in the supplementary materials.

Prediction 2: stronger crosslinguistic effects from the school language

In the second analysis, we investigate whether the crosslinguistic effect found above is stronger from the HL to the SL or from the SL to the HL. To this end, we included an interaction between 'previous other' and the language tested (HL vs. SL) in the model. This interaction was not significant, however ($\chi^2_{LR}(1) = 0.6, p = .42$). Thus, the prediction that crosslinguistic effects are stronger from the SL to the HL was not borne out. See the supplementary materials for the full model details.

Prediction 3: stronger crosslinguistic effects from French

In this third step, we investigate whether cross-linguistic effects are stronger between genealogically more closely related languages (French and Portuguese) than between more distantly related

languages (German and Portuguese). If literacy in French benefits more from literacy experience in Portuguese (or vice versa) than does literacy in German, we would expect an interaction between 'previous other' and language region (French- vs. German-speaking Switzerland). However, this interaction was not significant ($\chi^2(1) = 0.1, p = .93$).

In sum, we find no evidence for a difference between the effects between Portuguese on the one hand and French vs. German on the other hand. This is an unexpected finding since we had assumed that the genealogical proximity would be more helpful to Portuguese–French bilinguals. However, a closer examination of Figure 2 might suggest that the absence of such a difference could be due to differences in this effect between $T1 - T2$ and $T2 - T3$. Specifically, Figure 2 shows that the crosslinguistic effects between German and Portuguese may be stronger from $T2$ to $T3$ than from $T1$ to $T2$, and that an effect of typology may yet be present at $T1 - T2$. In order to further explore this possibility, we modelled the three-way interaction between 'previous other', language region, and time of data collection ($T2$ vs. $T3$). Compared to a model without this three-way interaction but with the lower-order interactions between 'previous other' and language region, 'previous other' and time, and language region and time, the model with the three-way interaction provided a better fit to the data ($\chi^2(1) = 10.2, p = .001$; see the supplementary materials for the full model details). The pattern suggested by this interaction is that the slope between 'previous other' and French is steeper between $T1 - T2$ than the one between 'previous other' and German, but that between $T2 - T3$, the opposite is true. In the aggregate, these effects cancelled each other out, yielding the null result reported above. This interaction was not replicated in the writing data.

Discussion

Using data from a longitudinal observational study with German–Portuguese and French–Portuguese bilingual school children, we set out to test three predictions that follow from Cummins' interdependence hypothesis and recent findings on second language acquisition: (1) There should be positive crosslinguistic longitudinal effects, even after taking into account within-language effects; (2) these effects should be stronger when they flow from the school language (SL; French or German), in which the children received most of their literacy training, to the heritage language (HL; Portuguese) than the other way around; (3) these effects should be stronger between the more closely related languages French and Portuguese than between the less closely related languages German and Portuguese.

As for the first prediction, the descriptive results presented in Figures 2 and 3 suggest that the children's performance in a given language at a given point in time can be estimated from their HL and SL scores from the year before. The regression model furthermore suggests that their score in a given language can more accurately be predicted if their earlier score in their other language are taken into account in addition to their previous score in the same language. This finding corroborates the first prediction we derived from the interdependence hypothesis (i.e. crosslinguistic longitudinal effects), but – as we will discuss below – it does not vindicate the assumption of interdependence.

As for the second prediction, we did not find any evidence that the participants' SL (French or German) contribute more to later literacy skills in their HL (Portuguese) than vice versa. Interpreting null results is always a challenge – after all, we cannot rule out that more fine-grained measures of literacy skills or a larger sample might have produced results more favourable to our second and third prediction. But the finding that crosslinguistic longitudinal effects are not clearly affected by the fact that the children received most of their literacy training in their SL does instil scepticism about the assumption of interdependence.

As for the third prediction, the evidence for a typological influence on crosslinguistic longitudinal effects is, at best, inconsistent. Specifically, we did not find stronger crosslinguistic effects between French and Portuguese than between German and Portuguese across the board. What we did find is a difficult-to-explain pattern that suggests that crosslinguistic effects in reading are relatively

strong between French and Portuguese at the onset but are weaker a year later, whereas these effects start out weak between German and Portuguese but grow stronger within a year. (Such an interaction was not observed for the writing data, see Vanhove and Berthele 2017).

Perhaps it would be possible to come up with a post-hoc explanation which assumes that Portuguese–French bilinguals had reached a lower threshold for more extensive HL–SL influence at $T1 - T2$ but had reached an upper threshold by $T2 - T3$, whereas Portuguese–German bilinguals had yet to reach the lower threshold at $T1 - T2$. However, already on conceptual grounds alone, we consider such an explanation problematic: without external theoretical justification, one can posit arbitrary double thresholds to salvage just about any prediction (see Takakuwa 2005). In addition to specifying where precisely these thresholds lie and why they exist, such a theoretical justification would also need to account for the absence of the three-way interaction for the other variables. In the meantime, we conclude that language genealogy does not exert a measurable influence on crosslinguistic longitudinal effects.

What are we to make of these results? We see two possible interpretations:

- (1) The crosslinguistic longitudinal effects show that Portuguese on the one hand and German or French on the other are interdependent, but HL-vs.-SL status and language genealogy do not affect interdependence.
- (2) The absence of effects of HL-vs.-SL status and language genealogy cast doubt on the usefulness of the interdependence hypothesis as a scientific framework, and the crosslinguistic longitudinal effects found can be explained in a different way.

Both of these interpretations are consistent with the data, but at present, we lean towards the second one for the following reasons. First, assuming that the bilinguals' languages are interdependent and that literacy skills acquired in one language may be applied in another, the prediction that the application of literacy skills mainly flows from the language in which these are predominantly acquired to the other language rather than vice versa seems obvious. Similarly, it seems evident that it is easier to apply a set of skills in a similar context (i.e. in a closely related language). Barring gross theoretical oversights on our part, two clear predictions based on the interdependence hypothesis were not borne out, casting doubt on the tenability of this hypothesis itself.

Second, as we pointed out in the introduction, the time lag between the different measurement times in this study, as well as in all other studies we are aware of, was chosen for practical but theoretically arbitrary reasons. However, this time lag must still be factored in when interpreting the results (see also Selig and Little 2012). If we take the significant crosslinguistic longitudinal effects at face value, then the interpretation must be that, at time T , Portuguese–French and Portuguese–German bilinguals applied some of their literacy subskills in Portuguese that they did not yet apply in French or German but that they did apply in the same French or German test a year later. Similarly, these bilinguals also applied some literacy subskills in French or German at time T that they did not apply in Portuguese until a year later. To us, such a face-value interpretation raises a number of questions:

- (1) What are these subskills that are applied in Portuguese but not in French/German until later?
- (2) What are the subskills that are applied in French/German but not in Portuguese until later?
- (3) Is it a coincidence that both kinds of subskills happen to be roughly equally important in reading and writing (as suggested by the lack of HL-vs.-SL effects for both skills)?
- (4) Why were these subskills not applied immediately (at time T) in the other language, seeing as they were useful in the same test at $T + 1$?
- (5) What, in the course of this one year, triggered the participants to apply these subskills in the other language?

These are five tough questions, but we suggest that the issue is with interpreting the crosslinguistic longitudinal effects for an arbitrary time lag at face value rather than with identifying some

unspecified literacy-related subskills. This brings us to the third reason why we favour explanation (2) above: the cross-linguistic longitudinal effects can be accounted for, at least in part, without assuming interdependence. The explanation concerns measurement error. As in most social science research, the variables in our (and other) studies are imperfect indicators of a construct. In most settings, random measurement error on the outcome variable does not bias the estimated parameters, though it does decrease the precision of the estimate. On the other hand, measurement error on predictor variables biases parameter estimates towards zero. One notable and underappreciated exception occurs when one of the predictor variables serves as a ‘statistical control’ in observational studies. Specifically, Brunner and Austin (2009) and Westfall and Yarkoni (2016) show that when two variables are correlated with each other and are both measured with error, and only one of them is actually causally related to a third variable, then measurement error will increase the probability that a significant effect is found between the unpredictable and the third variable, that is, the parameter estimate for the unpredictable variable is biased *away* from zero. This bias occurs even if the measurement errors themselves are unbiased. Moreover, larger samples do not alleviate the problem but compound it in that the unpredictable variable will be more likely to turn out significant.

The description above sounds familiar: Portuguese and French/German skills at a given point in time are likely to be correlated, and they are certain to be measured with error. Since Portuguese skills at time T are causally predictive of Portuguese skills at time $T + 1$, measurement error causes the causal relationship between French/German at time T and Portuguese at time $T + 1$ to be overstated – even after accounting for Portuguese at time T . This does not exclude the possibility that French/German at time T and Portuguese at time $T + 1$ really are causally related, but Brunner and Austin’s (2009) and Westfall and Yarkoni’s (2016) results compellingly show that these effects are at least overstated.²

Conclusions: post-mortem of a research project

The data analyzed here were collected with the objective of putting interdependence-related predictions to the test. During our research project, two kinds of doubts emerged: First, we came to realise that, as argued in this article, about any data pattern imaginable is consistent with some form of interdependence-based thinking, which impedes theoretical progress and does not help policymakers. Second, when analysing the data, we then came to realise that, if one wanted to put to the test the idea that investing in one language instrumentally benefits the other language, one would need to conduct randomised experiments – which we did not do either. In retrospect, we come to the conclusion that the interdependence hypothesis cannot fruitfully be tested in non-experimental studies. Moreover, this hypothesis, particularly if enriched by *post hoc* threshold explanations, makes rather trivial predictions, which can even easily be explained by mechanisms other than linguistic interdependence in the narrow sense: Cross-sectional correlations of measures across two or more languages within subjects may well be explained in terms of underlying linguistic proficiency, but they may just as well be due to the correlation of measurement errors or due to general, non-linguistic cognitive skills (working memory, general intelligence, world knowledge, and test-wiseness, to name but a few), whereas longitudinal correlations can at least be accounted for as the by-product of imperfectly measured variables. As long as scholars are unable to pinpoint proficiency threshold levels and other boundaries, such as the one between general cognitive capacities and the common underlying proficiency in bilinguals, interdependence-oriented research appears to be caught up in a never-ending circular process that does not contribute to the advancement of the theory.

There are different reasons and underlying values that lead scholars to advocate HL instruction or, more generally, bilingual education. In addition to the instrumental, transfer-oriented line of argumentation that is the object of this article, HL instruction can also be considered an important pillar of linguistic human rights (as proposed, e.g. in Skutnabb-Kangas, Phillipson, and Rannut 1994) or the aim of multilingual and multicultural modern societies (i.e. fostering HL proficiency

for its own sake). Whereas we do not question the importance of these latter values, our goal was to put to the test the instrumental idea that ‘investing’ in one language will have measurable beneficial effects in the other. Many studies, including ours, report crosslingual correlations between skills in different languages, but the question is not so much whether crosslingual correlations exists, but whether skills acquired in one language do or do not transfer into the other – crosslingual correlations need not be caused by such transfer. Moreover, scholars advocating bilingual education often cite meta-analyses that seem to suggest that bilingual programmes are efficient in promoting the academic achievement of minority students. However, in our view, it is premature to draw such conclusions from these meta-analyses: First, meta-analyses presuppose that effects of pedagogical programmes in rather diverse different settings involving completely different languages and language test formats can be standardised and compared. We do not think that it makes sense to standardise effects measured in various language tests across such different situations as Norwegian–Urdu and Catalan–Spanish bilingual education. Second, and more importantly, the gold standard for determining effects of two different educational paradigms that does not suffer from the usual biases due to self-selected samples in the bilingual education group would be experimental research involving the random assignment of participants to treatment groups. Unfortunately, such studies are a rare exception in the field: Out of the five (!) studies that remained in the meta-analysis by Reljić, Ferring, and Martin (2015), only one study (Hirst, Hannon, and Nutbrown 2010) randomly assigned the participants to the two different pedagogical treatment groups. Given the extremely small selection of studies in such analyses, and given that the studies retained generally do not allow causal inferences due to the methodological limitations, we cannot conclude that often small positive effects found in such meta-analyses provide robust evidence. Thus, as shown in our article as well as in other studies (e.g. Caprez-Krompák 2010; Moser, Bayer, and Tunger 2010), which were not taken into account in the meta-analysis by Reljić, Ferring, and Martin (2015), the empirical evidence for such instrumental effects of HL on SL proficiency is not overwhelming.

In addition, finding positive correlations between measurements in two languages is not sufficient for strong claims about interdependence – there are simply too many other explanations that are equally in line with such correlations. The empirically unsupported claims on strong interdependence and transfer effects raise the expectations regarding effects of HL instruction programmes, but such effects are weak at best. Expecting transfer miracles from heritage language instruction is therefore ultimately counterproductive for the legitimisation of heritage language instruction, since massive transfer is unlikely to happen and one of the main arguments in favour of this instruction can thus easily be shown to be void.

To be clear, we emphatically are not arguing that heritage language instruction should be abandoned; we merely think that citing the interdependence hypothesis as an argument in favour of heritage language instruction is counterproductive since this hypothesis is on shaky grounds both empirically and conceptually. But heritage language instruction can still be fostered for reasons cultural, social or aesthetic. However, such lines of argument belong to the social and political realms, whereas our goal in this article was to contribute to the investigation of linguistic transfer and interdependence phenomena in bi- and multilingual children with an immigration background. And in this respect, both our data and further conceptual considerations do not lend themselves to strong claims about beneficial interdependence effects.

Notes

1. Ardasheva, Tretter, and Kinny (2012) a priori operationalised the lower L2 threshold relative to a state-administered standardised test. They did not, however, test for the presence of a non-linear developmental slope with an inflection at this threshold.
2. Both Brunner and Austin (2009) and Westfall and Yarkoni (2016) suggest that researchers incorporate measurement errors in their statistical models, for example, using structural equation modeling (SEM). Such models have been used in interdependence studies and related research (e.g. Gebauer, Zaunbauer, and Möller 2013; Schoonen et al. 2011; Verhoeven 1994) but typically require multiple indicators for each construct, which the present project

does not have. Additionally, even when multiple indicators per construct are available, the causal relationships in SEMs may still be overstated: in longitudinal interdependence studies, the measurement errors are likely to be correlated both across languages and across different points in time inasmuch as the same or similar tasks are used. As Reddy (1992) demonstrates, ignoring such correlations may also cause the causal relationships in the SEM to be overstated. Unfortunately, estimating all these additional parameters would quickly exhaust the data available. In sum, we do not see any quick solutions to establish whether cross-linguistic longitudinal effects are due to interdependence (or another structural factor) or simply a by-product of measurement error. Resolving this issue is a clear desideratum for future studies.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding


This work was supported by Swiss Research Center on Multilingualism.

Notes on contributors

Raphael Berthele is Professor in multilingualism at the University of Fribourg. He directs the MA programmes in multi-lingualism studies and in foreign language didactics and he co-founded the Fribourg Institute of Multilingualism in 2008. His research interests cover different areas from cognitive to social aspects of multilingualism. During the last years, he has been focusing on the empirical investigation of receptive multilingualism and on convergence phenomena in the semantic and syntactic patterns in linguistic reference to space in multilinguals.

Jan Vanhove is a senior assistant at the Department of Multilingualism in Fribourg, Switzerland. He has done research on receptive multilingualism, crosslinguistic influence between closely related languages and lexical richness. He blogs semi-regularly about research design and statistics in applied linguistics and multilingualism research at <http://janhove.github.io>.

ORCID

Raphael Berthele  <http://orcid.org/0000-0003-1403-4116>

References

- Alderson, J. C. 1984. "Reading in a Foreign Language: A Reading Problem or a Language Problem?" In *Reading in a Foreign Language*, edited by J. C. Alderson and A. H. Urquhart, 1–27. London, UK: Longman.
- Ardasheva, Y., T. R. Tretter, and M. Kinny. 2012. "English Language Learners and Academic Achievement: Revisiting the Threshold Hypothesis." *Language Learning* 62 (3): 769–812. doi:10.1111/j.1467-9922.2011.00652.x.
- Baker, C. 2006. "Literacy, Biliteracy and Multiliteracies for Bilinguals." In *Foundations of Bilingual Education and Bilingualism*, edited by C. Baker, 320–345. Clevedon: Multilingual Matters.
- Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. doi:10.18637/jss.v067.i01.
- Beacco, J.-C., M. Byram, M. Cavalli, D. Coste, M. Egli Cuenat, F. Goullier, and J. Panthier. 2010. *Guide pour le développement et la mise en oeuvre de curriculums pour une éducation plurilingue et interculturelle*. Strasbourg: Conseil de l'Europe.
- Bishop, D. V. M., G. Laws, C. Adams, and C. F. Norbury. 2006. "High Heritability of Speech and Language Impairments in 6-Year-Old Twins Demonstrated Using Parent and Teacher Report." *Behavior Genetics* 36 (2): 173–184. doi:10.1007/s10519-005-9020-0.
- Brunner, J., and P. C. Austin. 2009. "Inflation of Type I Error Rate in Multiple Regression When Independent Variables are Measured With Error." *Canadian Journal of Statistics* 37 (1): 33–46. doi:10.1002/cjs.10004.
- Caprez-Krompák, E. 2010. *Entwicklung der erst- und zweitsprache im interkulturellen kontext*. Münster: Waxmann.
- Cárdenas-Hagan, E., C. D. Carlson, and S. D. Pollard-Durodola. 2007. "The Cross-Linguistic Transfer of Early Literacy Skills: The Role of Initial L1 and L2 Skills and Language of Instruction." *Language Speech and Hearing Services in Schools* 38: 249. doi:10.1044/0161-1461(2007)026.
- Collins, L. M. 2006. "Analysis of Longitudinal Data: The Integration of Theoretical Model, Temporal Design, and Statistical Model." *Annual Review of Psychology* 57: 505–528. doi:10.1146/annurev.psych.57.102904.190146.

- Collins, L. M., and J. W. Graham. 2002. "The Effect of the Timing and Spacing of Observations in Longitudinal Studies of Tobacco and Other Drug Use: Temporal Design Considerations." *Drug and Alcohol Dependence* 68 (Suppl. 1): 85–96. doi:10.1016/S0376-8716(02)00217-X.
- Cummins, J. 1976. *The Influence of Bilingualism on Cognitive Growth: A Synthesis of Research Findings and Explanatory Hypotheses* (Working Papers on Bilingualism No. 9). Toronto: The Ontario Institute for Studies in Education.
- Cummins, J. 1979. "Linguistic Interdependence and the Educational Development of Bilingual Children." *Review of Educational Research* 49 (2): 222–251. doi:10.3102/00346543049002222.
- Cummins, J. 1996. *Negotiating Identities: Education for Empowerment in a Diverse Society*. Los Angeles: California Association for Bilingual Education.
- Cummins, J. 2001. *Negotiating Identities: Education for the Empowerment in a Diverse Society*. 2nd ed. Los Angeles: California Association for Bilingual Education.
- Cummins, J. 2009. "Pedagogies of Choice: Challenging Coercive Relations of Power in Classrooms and Communities." *International Journal of Bilingual Education and Bilingualism* 12 (3): 261–271. doi:10.1080/13670050903003751.
- Cummins, J. 2017. "Teaching for Transfer in Multilingual Educational Contexts." In *Encyclopedia of Language and Education, 3rd Edition, Volume 5: Bilingual Education*, edited by O. García, and A. Lin, 103–115. New York: Springer.
- Da Fontoura, H. A., and L. S. Siegel. 1995. "Reading, Syntactic, and Working Memory Skills of Bilingual Portuguese-English Canadian Children." *Reading and Writing: An Interdisciplinary Journal* 7: 139–153.
- Desgrippes, M., and A. Lambelet. 2017. "On the Sociolinguistic Embedding of Portuguese Heritage Language Speakers in Switzerland: Socioeconomic Status and Home Literacy Environment (HELASCOT Project)." In *Heritage and School Language Literacy Development in Migrant Children: Interdependence or Independence?* edited by R. Berthele, and A. Lambelet, 34–57. Clevedon: Multilingual Matters.
- Dressler, C., and M. L. Kamil. 2006. "First- and Second-Language Literacy." In *Developing Literacy in Second Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*, edited by D. August, and T. Shanahan, 197–238. Washington DC.
- Esser, H. 2009. "Der Streit um die Zweisprachigkeit: Was bringt die Bilingualität?" In *Streitfall Zweisprachigkeit = The bilingualism controversy*, edited by I. Gogolin and U. Neumann, 69–88. Wiesbaden: VS Verlag für Sozialwissenschaften.
- García, G. E., R. T. Jiménez, and P. D. Pearson. 1998. "Metacognition, Childhood Bilingualism, and Reading." In *Metacognition in Educational Theory and Practice*, edited by D. J. Hacker, J. Dunlosky, and A. C. Graesser, 193–219. New York: Routledge.
- Gebauer, S. K., A. C. M. Zaunbauer, and J. Möller. 2013. "Cross-Language Transfer in English Immersion Programs in Germany: Reading Comprehension and Reading Fluency." *Contemporary Educational Psychology* 38 (1): 64–74. doi:10.1016/j.cedpsych.2012.09.002.
- Hall, J. K., A. Cheng, and M. T. Carlson. 2006. "Reconceptualizing Multicompetence as a Theory of Language Knowledge." *Applied Linguistics* 27 (2): 220–240. doi:10.1093/applin/aml013.
- Hirst, K., P. Hannon, and C. Nutbrown. 2010. "Effects of a Preschool Bilingual Family Literacy Programme." *Journal of Early Childhood Literacy* 10 (2): 183–208. doi:10.1177/1468798410363838.
- Hulstijn, J. H. 2015. *Language Proficiency in Native and Non-Native Speakers: Theory and Research*. John Benjamins. doi:10.1075/llt.41.
- Idiazabal, I., and L. M. Larrigan. 1997. "Transfert de maîtrises discursives dans un programme d'enseignement bilingue basque-espagnol." *Acquisition et Interaction En Langue Étrangère* 10: 107–125.
- Jarvis, S. 2000. "Methodological Rigor in the Study of Transfer: Identifying L1 Influence in The Interlanguage Lexicon." *Language Learning* 50 (2): 245–309. doi:10.1111/0023-8333.00118.
- Jarvis, S. 2010. "Comparison-Based and Detection-Based Approaches to Transfer Research." *EUROSLA Yearbook* 10: 169–192. doi:10.1075/eurosla.10.10jar.
- Jarvis, S. 2012. "The Detection-Based Approach: An Overview." In edited by S. Jarvis, and S. A. Crossley, 1–33. Bristol: Multilingual Matters.
- Jarvis, S., and S. A. Crossley, eds. 2012. *Approaching Language Transfer Through Text Classification: Explorations in the Detection-Based Approach*. Bristol: Multilingual Matters.
- Jarvis, S., and A. Pavlenko. 2008. *Crosslinguistic Influence in Language and Cognition*. New York: Routledge.
- Kecskes, I., and T. Papp. 2003. "How to Demonstrate the Conceptual Effect of L2 on L1? Methods and Techniques." In *Effects of the Second Language on the First*, edited by V. Cook, 247–267. Clevedon: Multilingual Matters.
- Lambelet, A., R. Berthele, M. Desgrippes, C. Pestana, and J. Vanhove. 2017. "Testing Interdependence in Portuguese Heritage Speakers in Switzerland: The HELASCOT Project." In *Heritage and School Language Literacy Development in Migrant Children: Interdependence or Independence?* edited by R. Berthele, and A. Lambelet, 26–33. Bristol: Multilingual Matters.
- Lambelet, A., M. Desgrippes, and J. Vanhove. 2017. "The Development of Argumentative and Narrative Writing Skills in Portuguese Heritage Speakers in Switzerland (HELASCOT Project)." In *Heritage and School Language Literacy Development in Migrant Children: Interdependence or Independence?* edited by R. Berthele, and A. Lambelet, 83–96. Bristol: Multilingual Matters.
- Lee, J.-W., and D. L. Schallert. 1997. "The Relative Contribution of L2 Language Proficiency and L1 Reading Ability to L2 Reading Performance: A Test of the Threshold Hypothesis in an EFL Context." *TESOL Quarterly* 31: 713–739.

- Moser, U., N. Bayer, and V. Tunger. 2010. "Erstsprachförderung bei Migrantenkindern in Kindergärten." *Zeitschrift Für Erziehungswissenschaft* 13 (4): 631–648. doi:10.1007/s11618-010-0144-5.
- Muñiz-Swicegood, M. 1994. "The Effects of Metacognitive Reading Strategy Training on the Reading Performance and Student Reading Analysis Strategies of Third Grade Bilingual Students." *Bilingual Research Journal* 18 (1-2): 83–97. doi:10.1080/15235882.1994.10162659.
- Odlin, T. 1989. *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.
- Pestana, C., A. Lambelet, and J. Vanhove. 2017. "Reading Comprehension in Portuguese Heritage Speakers in Switzerland (HELASCOT Project)." In *Heritage and School Language Literacy Development in Migrant Children: Interdependence or Independence?* edited by R. Berthele, and A. Lambelet, 58–82. Bristol: Multilingual Matters.
- Proctor, C. P., D. August, C. Snow, and C. D. Barr. 2010. "The Interdependence Continuum: A Perspective on the Nature of Spanish–English Bilingual Reading Comprehension." *Bilingual Research Journal* 33 (1): 5–20. doi:10.1080/15235881003733209.
- R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <http://www.r-project.org/>.
- Reddy, S. K. 1992. "Effects of Ignoring Correlated Measurement Error in Structural Equation Models." *Educational and Psychological Measurement* 52 (3): 549–570. doi:10.1177/0013164492052003005.
- Reljić, G., D. Ferring, and R. Martin. 2015. "A Meta-Analysis on the Effectiveness of Bilingual Programs in Europe." *Review of Educational Research* 85 (1): 92–128. doi:10.3102/0034654314548514.
- Rimfeld, K., P. S. Dale, and R. Plomin. 2015. "How Specific is Second Language-Learning Ability? A Twin Study Exploring the Contributions of First Language Achievement and Intelligence to Second Language Achievement." *Translational Psychiatry* 5: e638. doi:10.1038/tp.2015.128.
- Rolstad, K., K. S. Mahoney, and G. V. Glass. 2005a. "Weighing the Evidence: A Meta-Analysis of Bilingual Education in Arizona." *Bilingual Research Journal* 29 (1): 43–67. doi:10.1080/15235882.2005.10162823.
- Rolstad, K., K. Mahoney, and G. V. Glass. 2005b. "The Big Picture: A Meta-Analysis of Program Effectiveness Research on English Language Learners." *Educational Policy* 19 (4): 572–594. doi:10.1177/0895904805278067.
- Schepens, J. 2015. "Bridging Linguistic Gaps: The Effects of Linguistic Distance on the Adult Learnability of Dutch as an Additional Language." *PhD thesis*, Radboud University, Nijmegen.
- Schoonen, R., J. Hulstijn, and B. Bossers. 1998. "Metacognitive and Language-Specific Knowledge in Native and Foreign Language Reading Comprehension: An Empirical Study among Dutch Students in Grades 6, 8 and 10." *Language Learning* 48 (1): 71–106. doi:10.1111/1467-9922.00033.
- Schoonen, R., A. van Gelderen, R. D. Stoel, J. Hulstijn, and K. de Glopper. 2011. "Modeling the Development of L1 and EFL Writing Proficiency of Secondary School Students." *Language Learning* 61 (1): 31–79. doi:10.1111/j.1467-9922.2010.00590.x.
- Selig, J. P., and T. D. Little. 2012. "Autoregressive and Cross-Lagged Panel Analysis for Longitudinal Data." In *Handbook of Developmental Research Methods*, edited by B. Laursen, T. D. Little, and N. A. Card, 265–278. New York, NY: Guilford.
- Shuy, R. 1978. "Problems in Assessing Language Ability in Bilingual Education Programs." In *Bilingual Education*, edited by H. Lafontaine, H. Persky, and L. Golubchick, 376–381. Wayne, NJ: Avery.
- Singley, M. K., and J. R. Anderson. 1988. "A Keystroke Analysis of Learning and Transfer in Text Editing." *Human-Computer Interaction* 3: 233–274.
- Skutnabb-Kangas, T., R. Phillipson, and M. Rannut. 1994. *Linguistic Human Rights: Overcoming Linguistic Discrimination*. Berlin: Mouton de Gruyter.
- Stromswold, K. 2001. "The Heritability of Language: A Review and Metaanalysis of Twin, Adoption, and Linkage Studies." *Language* 77 (4): 647–723.
- Takakuwa, M. 2005. "Lessons from a Paradoxical Hypothesis: A Methodological Critique of the Threshold Hypothesis." In *ISB4: Proceedings of the 4th International Symposium on Bilingualism*, edited by J. Cohen, K. T. McAlister, K. Rolstad, and J. MacSwan, 2222–2232. Somerville, MA: Cascadia.
- van Gelderen, A., R. Schoonen, R. D. Stoel, K. de Glopper, and J. Hulstijn. 2007. "Development of Adolescent Reading Comprehension in Language 1 and Language 2: A Longitudinal Analysis of Constituent Components." *Journal of Educational Psychology* 99 (3): 477–491. doi:10.1037/0022-0663.99.3.477.
- Vanhove, Jan. 2013. "The Critical Period Hypothesis in Second Language Acquisition: A Statistical Critique and a Reanalysis." *PLOS ONE* 8 (7): e69172. <http://dx.doi.org/10.1371/journal.pone.0069172>.
- Vanhove, Jan. 2014.
- Vanhove, J., and R. Berthele. 2017. "Testing the Interdependence of Languages (HELASCOT Project)." In *Heritage and School Language Literacy Development in Migrant Children: Interdependence or Independence?* edited by R. Berthele, and A. Lambelet, 97–118. Bristol: Multilingual Matters.
- Vanhove, J. The Critical Period Hypothesis in Second Language Acquisition: A Statistical Critique and a Reanalysis." *PLOS ONE* 9 (7): e102922.
- Verhoeven, L. T. 1994. "Transfer in Bilingual Development: The Linguistic Interdependence Hypothesis Revisited." *Language Learning* 44 (3): 381–415. doi:10.1111/j.1467-1770.1994.tb01112.x.
- Weinreich, U. 1953. *Languages in Contact: Findings and Problems*. New York: Linguistic Circle of New York.
- Westfall, J., and T. Yarkoni. 2016. "Statistically Controlling for Confounding Constructs is Harder Than You Think." *PLOS ONE* 11 (3): e0152719. doi:10.1371/journal.pone.0152719.