

# **The influence of code-mixing and speaker information on perception and assessment of foreign language proficiency: An experimental study**

**Raphael Berthele**

University of Fribourg, Switzerland

## **Abstract**

The study draws on different lines of research on the influence of social and other information on the evaluation of language production in school contexts. On the one hand, names or other background information is well known to influence teachers and other gatekeepers' evaluations, and on the other hand, code-switching and other non-standard features in pupils' language production are also known to affect assessment outcomes not only of linguistic skills but also of general academic potential. Taking into account these two research traditions, this study investigates the influence of different ethnically marked names and code-switches on teachers' evaluations of pupils' oral proficiency in French as a foreign language. Three authentic oral texts were rerecorded once by inserting German words and once without such inserts. Additionally, these samples were presented either as stemming from a bilingual Swiss German native or from a multilingual Swiss-German Serbian boy. A total of 157 future teachers rated the speech samples with respect to different dimensions (fluency, correctness, but also the pupil's academic potential in general). The analyses provide evidence for positive and negative stereotyping of the Serbian first name, and there is also an unexpected interaction with code-mixing into German: without insertional mixing, the texts with a Balkan name are perceived as being superior, but with such mixing this superiority is lost and turns into significantly lower assessment scores.

## **Keywords**

code-mixing, code-switching, education, foreign language learning, migration, social inequalities, stereotyping

## **1 Context and background**

Bilingualism and multilingualism and their overt manifestations are constantly subject to evaluative judgments both within and outside of linguistics. In linguistics, the focus has long since moved away from the stereotypically negative assessment of bilingualism in general and interference phenomena in particular (Adler, 1977; Weisgerber, 1966) to a differentiated description of bilingualism and multilingualism as a specific form of competence (Grosjean, 1985; Herdina & Jessner, 2002; Muysken, 2000; Li Wei, 2004). Oftentimes bi- and multilingualism are even celebrated as being not only the normal, but even a more noble form of

linguistic competence, either given its deemed linguistic and acquisitional (Berthele, 2011; Cenoz, 2003; Jessner, 2000), its cognitive (Bialystok, 1987; Chertkow, et al., 2010; Diaz & Klinger, 1991) or its intercultural potentials (Byram & Hu, 2009). During the last decades, the 'language regard' (Preston, 2010) for bi- and multilingualism of scholarly research has thus evolved, reconsidering anew the pioneering study by Weinreich (1953) and thus developing a genuine scholarly interest in interferences resulting from contact to a more general interest in bi- and multidirectional transfer phenomena (Pavlenko & Jarvis, 2002), code-switching and code-mixing (Auer, 1999; Gardner-Chloros, 2009; Muysken, 2000; Poplack, 1980) and convergence (Auer & di Luzio, 1988; Backus, 2004). Given this interest in the investigation of bilingualism as a normal and natural form of language usage and proficiency, many actors in the educational domain, particularly in second and foreign language learning and teaching research, strive for a new view on multilingual competence and multilingual practices in the (language learning) classroom (for recent examples see Bono & Melo-Pfeifer, 2010; Butzkamm & Caldwell, 2009). More specifically, at least in some western European countries, and stimulated by the work of the language policy division of the Council of Europe (Trim & North, 2003), there is a perceptible and sustained endeavor for a new regard for bi- and multilingual competence, for its peculiarities and for its potential in classrooms and in the school curriculum. The considerable proportion of multilingual minority students with a migration background in western European classrooms and the questions of heritage language maintenance or linguistic assimilation and of the educational success of labor migrant children have also led to new approaches to linguistic and cultural diversity in the educational domain.

Even though the scientific regard for bilingualism has thus evolved considerably during the last decades, their effect on educators' and policy makers' attitudes and folk theories of bi- or multilingualism has often been modest to absent. Even if classes on language acquisition and bi- and multilingualism are gradually entered into the curricula of future teachers, it would be quite exaggerated to claim that there might emerge a generalized school culture that takes into account pluri- and multilingual competence phenomena, be it in the teaching and learning setups or be it in the assessment of plurilingual proficiency and skills (see Lenz & Berthele, 2010 for suggestions in this respect).

The study reported here addresses the question of how manifestations of bilingual proficiency are regarded and assessed by future primary school teachers. The experimental study draws on two lines of research. First, studies on the assessment of ethnically accented or bilingually marked speech style show that teachers, above all if they have no or little knowledge of bilingualism and bilingual practices, negatively evaluate such markers in pupils' speech (Greene & Walker, 2004; Hughes, Shaunessy & Brice, 2006; Ramirez & Milk, 1986; Rubin, 1992). As discussed in Hughes et al. (2006), code-switching can be considered as a marker of lack of proficiency (e.g., lexical gaps in one language filled by words from the other) or as a manifestation of high bilingual competence, since especially complex

switching (Muysken, 2006) requires the bilingual individual to master not only one but two grammars. Moreover, one can also argue that insertional code-mixing represents a valuable achievement strategy of second-language learners (cf. Faerch & Kasper, 1983, p. 46).

On the other hand, there is a considerable body of research from different disciplines that shows how the categorization and the assessment of all kinds of (tacitly assumed or actually performed) skills and potentials of people are influenced by ethnic, social or other information. In perceptual dialectology, Niedzielski (1999) has shown how information about the origin of speakers influences the perception of phonological parameters. The influence of ethnically marked names in submitted CVs on callback rates has been shown in many studies in very different countries (see e.g., Rebzani, 2005 for the French context; Booth, Leigh & Varganova, 2009 for the Australian context). Ethnicity also influences the rapidity of assessment in soccer challenges and housing discrimination (Purnell, Idsardi & Baugh, 1999; Wagner-Egger, Gygax, & Ribordy, submitted). Finally, the influence of teacher expectations on students' performances is a robust finding in educational research since Rosenthal and Jacobson's classical study (Jussim & Harber, 2005; Rosenthal & Jacobson, 1992).

## 2 Research questions

The study discussed in the remainder of this article draws on both lines of research discussed in section 1, and aims at investigating a process that is deemed to be at the very center of the educational challenges referred to earlier. The main research question asked is how teachers assess overt manifestations of bilingual competence in the form of code-mixing into a language that is not the target language of the foreign language classroom in question. Moreover, the study investigates the interaction between the presence or absence of such code-mixing with ethnic and language biographical information of the pupils who are being assessed. The experimental design of the study will be described later. The main hypotheses investigated are the following:

1. The same oral text in a foreign language will be assessed differently according to the label with which it is presented to the assessors. The expected direction of the effect is that Swiss name labels will lead to more positive evaluations than immigrant name labels. I refer to this effect as the name stereotype effect.
2. Violations of monolingual speech in the foreign language classroom – in the form of insertional code-mixing in the sense of Muysken (2006) – are perceived and negatively evaluated by teachers. This effect will be called the *mixing depreciation effect* and it is expected to act not only on the assessment of proficiency underlying the particular oral production – where they make sense, at least from a normative point of view – but also to other

types of assessment such as the expectations regarding the academic potential of the student.

3. There is an interaction between the *mixing depreciation effect* and the *name stereotype effect*, that is to say, the assessors are expected to treat the code-mixing variable differently according to the name of the student. There is no *a priori* expectation regarding the direction of this effect, since one can hypothesize that code-mixing confirms stereotypical expectations of language proficiency in migrant children which could either lead to a relatively positive assessment ('the student mixes languages, which is normal in his case, but otherwise the production is quite acceptable') or to a more negative assessment ('the student, given his status as a bilingual migrant, is over-challenged by the additional foreign language, therefore this production is bad, as expected').

### **3 Method**

#### ***3.1 Participants***

The participants in this experiment were 157 future primary education teachers, all students at the bilingual Haute École Pédagogique Fribourg/Pädagogische Hochschule Freiburg in Switzerland (henceforth HEPFR). The students were either in the first or second year of their three-year curriculum. The subjects were purposely misinformed that the researchers were working on educational standards for Swiss foreign language classrooms and that they needed the students' help to calibrate authentic oral productions of Swiss German primary school pupils learning French. The field experiment took place during regular classes in the usual institutional environment. The students were distributed a form on which they had to give their assessment of different dimensions (see later) of three oral stimuli that were presented to them one after the other.

#### ***3.2 Stimuli***

The basis for the stimuli was three authentic oral learner productions in French as a foreign language. These oral texts had been collected in the context of a large-scale empirical investigation that actually served the purpose that had misleadingly been presented to the subjects, that is, the standard setting for foreign language education in Swiss primary schools (see Lenz, 2007). The three productions were responses to three typical tasks in the primary school foreign language classroom: the short description of a day in the pupil's life ('parler d'une journée'), the description of the pupil's class ('portrait de classe'), and a description of the pupil's town ('ta ville'). For the purposes of this study, three productions judged 'typical' for sixth graders by the authors of the aforementioned study (Lenz, 2007) have been transcribed and slightly modified (mostly shortened). For the condition with inserts, three lexical items were translated into German (see example 1), in order to

simulate the learners' word finding problems and the insertion of a native language word as a bilingual strategy to resolve the problem.

(1) Stimuli: example (partial transcription of the orally presented text 'ma classe')

+INSERTS: [...] Je sors de la maison et je vais à gauche et d'abord je vais GERADEAUS et ensuite je suis à l'école. Et je vais avec Ruben à l'école. ALSO elle vient ABHOLEN moi et ensuite nous allons aller à l'école. [...]

–INSERTS: [...] Je sors de la maison et je vais à gauche et d'abord je vais tout droit et ensuite je suis à l'école. Et je vais avec Ruben à l'école. Alors elle vient chercher moi et ensuite nous allons aller à l'école. [...]

I go out of the house and I turn left and first I go straight ahead and then I am at my school. I walk to school with ruben. so she comes to pick me up and then we go to school.

The stimuli were rerecorded by three female Swiss German volunteers who all speak excellent French but who deliberately exaggerated the Swiss German accent in French to make the recordings credible. Each text was recorded twice, once with the three inserts (henceforth *+inserts condition*), once without them (henceforth *–inserts condition*). The recordings are between 1 and 2 minutes long. They have been resynthesized using the praat phonetics software (version 4.1.44) in order to simulate a young boy's voice (mean F0 set to 250 Hz, all formants shifted with a 1.1 ratio). The resynthesized stimuli were presented to four uninformed test persons, asking them to guess the age of the person. All respondents judged the speakers to be between 8 and 13 years old, which is about the appropriate range. No one, neither in the pretesting nor in the testing context, manifested overt suspicion regarding the authenticity of the learner language stimuli.

**Table 1.** Participant groups and stimulus presentation.<sup>1</sup>

Group	Text	Label	Inserts
A	Ta ville	Dragan, 12 years, sixth grade, native language: Serbian	+inserts
	Parler d'une journée	Luca, 11 years, fifth grade, native language: Swiss German	–inserts
	Portrait de classe	Goran, 10 years, fifth grade, native language: Serbian	+inserts
B	Ta ville	Luca, 12 years, sixth grade, native language: Swiss German	+inserts
	Parler d'une journée	Dragan, 11 years, fifth grade, native language: Serbian	–inserts
	Portrait de classe	Tim, 10 years, fifth grade, native language: Swiss German	+inserts

C	Ta ville	Dragan, 12 years, sixth grade, native language: Serbian	-inserts
	Parler d'une journée	Luca, 11 years, fifth grade, native language: Swiss German	+inserts
	Portrait de classe	Goran, 10 years, fifth grade, native language: Serbian	-inserts
D	Ta ville	Luca, 12 years, sixth grade, native language: Swiss German	-inserts
	Parler d'une journée	Dragan, 11 years, fifth grade, native language: Serbian	+inserts
	Portrait de classe	Tim, 10 years, fifth grade, native language: Swiss German	-inserts

### 3.3 Procedure

The participants were divided in four groups (see Table 1). Each group was administered three evaluation subtasks, each based on a stimulus text. The three stimuli were presented in a fixed order to the participants, but with variation regarding the inserts and the ethnic and linguistic identity of the speaker (Table 1). The participants always saw a slide with short information on the task the pupils had been asked to carry out, together with the information on the (invented) ethnic and linguistic identity of the speaker (see Figure 1).

The names Dragan and Goran, together with the indication that the first language is Serbian, are supposed to trigger a possible 'Balkan stereotype'; the names Luca and Tim are common first names in German-speaking Switzerland and are supposed to trigger the 'Swiss stereotype'. The labels were presented visually and the same information was read out loud by the field workers before the audio file was played.

<p>Texte 3 – Portrait de classe</p>	<p>Goran 10 ans, langue maternelle serbe</p>	<p>Text 3 – Portrait of your class. Goran, 10 years old, native language Serbian.</p> <p>Situation: A partner class from Zurich would like to know more about your school and your class. Thus you send them a video message.</p> <p>Task: Present your class. Tell something about each of the following points: Class, number of students, number of girls and boys, subjects: which ones you like/dislike, schoolhouse, your way to school.</p>
<p>Situation:</p> <p>Une classe jumelée de Zurich aimerait mieux connaître ton école et ta classe. C'est pourquoi vous leur envoyez un message vidéo.</p> <p>Devoir:</p> <ul style="list-style-type: none"> <li>• Présente ta propre classe.</li> <li>• Dis quelque chose sur chacun des points suivants:</li> <li>• Classe, nombre d'élèves, nombre de filles, nombre de garçons, matières scolaires: lesquelles 😊, lesquelles ☹️, bâtiment scolaire, chemin de l'école</li> </ul>		

**Figure 1.** Visually presented context information for text 3 (Balkan condition).

The participants had to assess the oral production examples according to the following dimensions (all forms were in French, one of the two official languages of the HEPFR): vocabulary richness, pronunciation, fluidity, grammatical correctness, originality, intelligibility, influence of other languages (interferences), global evaluation of text, global proficiency level of speaker, academic potential of speaker. Only two participants refused to assess the last dimension. All assessments were done via a 6-point Likert scale, ranging from very low to very high. The 'negative' dimension of influence of other languages has been recoded in order to have high values represent small amounts of interference and vice versa.

#### **4 Data analysis and results**

The data sets obtained via the procedures described in the last section were analyzed comparing the four different conditions. In a first step, standardized Z-scores of the dependent variables were generated within each of the three oral texts individually (but neglecting the +inserts and -inserts conditions) to cancel the potentially biasing effect of differing numbers of assessments in each cell on the means. In a second step, the standardized results of the assessment variables were reduced to five dimensions, in order to minimize the problems of multiple testing (inflation of family-wise type I error rate). A global *correctness* variable was generated by calculating the mean scores of the pronunciation and the grammatical correctness variable. A *communicative richness* variable was generated by calculating the means of the fluidity, intelligibility, vocabulary richness and originality variables. A *global assessment* variable represents the means of the *global evaluation* of the text and the *global language proficiency evaluation* variables. The two dimensions of academic potential and interferences are based on one Likert item respectively.

Based on these data transformations, descriptive and inferential statistical analyses were carried out in order to describe the variance in the data when comparing the four conditions (see Figure 2) and to test the hypotheses formulated earlier.

##### ***4.1 Descriptive statistics comparing the four presentation conditions***

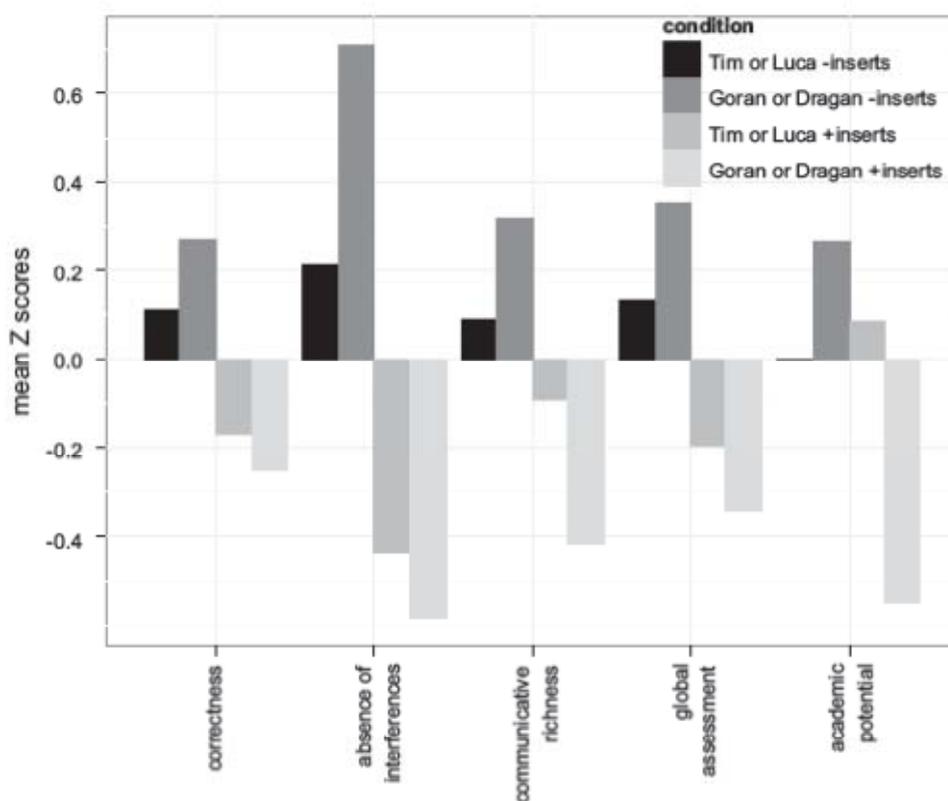
Figure 2 suggests that the ethnic label attached to the oral text production indeed influences the outcome of the assessment. However, as shown in Figure 2, the direction of the name stereotype effect is surprising in different respects (cf. hypothesis 1 given earlier).

The general tendency for the five dependent variables across the four conditions seems to be the same: the indigenous label coincides consistently with assessments closer to the average (the 0 line) than the Balkan label. Moreover, the influence of the inserts leads, as expected, to more negative assessments compared to the – inserts condition. There is only one exception to this tendency: the mean assessment of the academic potential of the pupil is higher for the Tim/Luca

condition with inserts than for Tim/Luca without inserts. In the remainder of this analysis, we will test which of these differences are statistically significant.

#### ***4.2 Testing the influence of the factors and their interaction***

A multivariate GLM analysis with the insertion variable as the first and the name as the second factor and the five assessment variables as dependent variables was carried out. The goal was to test the three hypotheses formulated at the beginning of this article. Hypotheses 1 and 2 correspond to the main effects of the two factors on the dependent variables. Hypothesis 3 was tested by looking at interactions between the two factors. Moreover, this analysis allows assessing the impact of the factors on each of the five assessment dimensions separately.



**Figure 2.** Mean results for the five dependent variables (Z-scores) across four conditions.

The name stereotype effect was significant, although the effect size is relatively small (see Table 2). We can thus conclude that, indeed, the pupil's name does have an influence on the assessment scores. However, this influence is not straightforward, as other results from research discussed earlier seem to suggest. We will discuss the role of the name factor in more detail later.

There also was a highly significant mixing depreciation effect with a much larger effect size (Table 2). Along the lines of the classical research on the perception and

assessment of code-switching and other stylistically marked features of student speech (Ramirez & Milk, 1986), we can clearly state that the violation of the monolingual norm also gets sanctioned in a similar way in the foreign language learning classroom, thus confirming hypothesis 2 regarding the mixing depreciation effect.

Finally, there is a medium effect (Table 2) of the interaction between the name and the mixing variables. At first sight, all three hypotheses can be confirmed: The variation of the name label has an influence on the assessment of the pupil's text productions, and even more so does the presence of insertional mixing influence the assessment. And finally, the two factors interact, in other words, the role of the name is not the same in presence or absence of mixing (or vice versa).

**Table 2.** Multivariate analyses of the two factors and their interaction. Effect sizes are assessed by calculating partial Eta squares.

Pillai's Trace	Value	<i>F</i>	df	<i>p</i>	effect size
factor			hypothesis, error		
Mixing depreciation effect	0.264	32.5	5,453	< 0.001**	0.264
Name stereotype effect	0.030	2.9	5,435	= 0.015*	0.030
Interaction mixing × name	0.077	7.5	5,453	< 0.001**	0.077

\*  $0.01 < p < .05$   
 \*\*  $p < .01$

**Table 3.** Univariate analyses for the five dependent variables. Effect sizes are assessed by calculating partial Eta squares.

Dependent variable	Tests of between-subjects effects				effect size
	df	Mean square	<i>F</i>	<i>p</i>	
<b>Mixing depreciation effect</b>					
<b>Correctness</b>	1	17.4	31.3	< 0.001**	0.064
<b>Cross-linguistic interferences</b>	1	101.7	144.1	< 0.001**	0.240
<b>Communicative richness</b>	1	22.1	43.8	< 0.001**	0.088
<b>Global assessment</b>	1	28.8	48.6	< 0.001**	0.096
<b>Academic potential</b>	1	14.3	15.4	< 0.001**	0.033
<b>Name stereotype effect</b>					

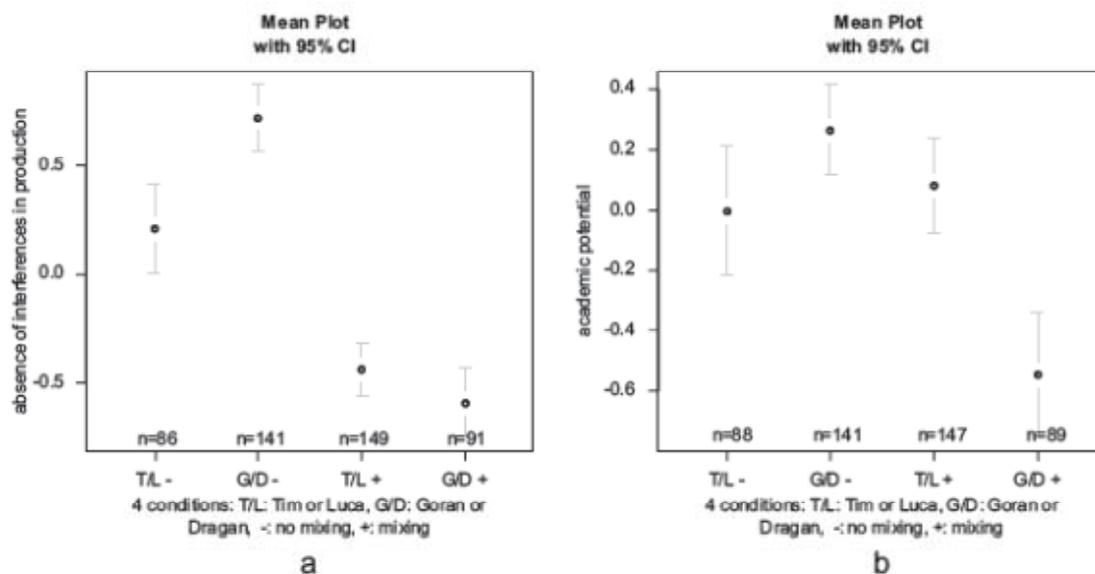
<b>Correctness</b>	1	0.2	0.28	= 0.599	0.001
<b>Cross-linguistic interferences</b>	1	3.3	4.7	= 0.030*	0.010
<b>Communicative richness</b>	1	0.3	0.5	= 0.481	0.001
<b>Global assessment</b>	1	0.2	0.3	= 0.578	0.001
<b>Academic potential</b>	1	3.6	3.9	= 0.048*	0.009
<b>Name s. × mixing d.</b>					
<b>Correctness</b>	1	1.6	2.8	= 0.096	0.006
<b>Cross-linguistic interferences</b>	1	11.5	16.3	< 0.001**	0.034
<b>Communicative richness</b>	1	8.0	15.9	< 0.001**	0.034
<b>Global assessment</b>	1	3.4	5.8	= 0.016*	0.013
<b>Academic potential</b>	1	21.3	23.0	< 0.001**	0.048
* $0.01 < p < .05$					
** $p < .01$					

#### ***4.3 Differences across the dependent variables in more detail***

As Figure 2 shows, not all of the dependent variables are affected by the factors with the same magnitude. In order to assess the differences across the five variables the results of the GLM between-subjects effects can be used (see Table 3 for a summary).

These results reveal that the name variable directly affects the assessment of cross-linguistic interferences (see Figure 3a) and the assessment of the academic potential (see Figure 3b). Interestingly, they are not affected in the same way: the Balkan name label, with respect to the mean of the + and -inserts conditions, yields assessments of relatively low amounts of interference

(i.e. high scores on the recoded Likert scale: mean difference Tim/Luca – Goran/Dragan of -0.176,  $p < 0.03$  with Bonferroni correction). Thus, we find a name stereotype effect in the unexpected direction, with the Balkan name triggering comparatively good assessments of the learners' production. However, in the case of academic potential, the tendency is inverted, with a mean difference of 0.183, that is, with higher academic potential for the Tim/Luca condition (Bonferroni corrected  $p < 0.048$ , again for the mean of the + and -inserts conditions). These at least partially surprising findings will be discussed and interpreted in the concluding part of this article. None of the other dependent variables show a main effect of the name label.



**Figure 3.** Means and 95% confidence intervals for the assessment of cross-linguistic interferences (a) and academic potential (b) across the four conditions. The higher the values in (a), the less interference is diagnosed by the raters.

The presence or absence of insertional mixing has an important impact on all five variables. We find a large effect for the cross-linguistic interferences (see Table 3 for statistical details), medium to large effects for global assessment of proficiency and communicative richness, and a small effect on the assessment of academic potential. A mixing depreciation effect can thus be reported for all dimensions of assessment, including the pupil's academic potential, a dimension that cannot be said to be directly and intrinsically related to the presence or absence of insertional mixing (see Section 5).

Hypothesis 3, regarding the interaction of the two factors, is confirmed by the between-subjects effects: four out of the five dependent variables yielded significant effects. A medium effect was found for the academic potential (see Table 3), small to medium effects were found for the assessment of cross-linguistic interferences and of communicative richness, and a small effect was found for the global assessment of the pupil's proficiency. And finally, there is a marginally significant effect for the fifth variable, correctness.<sup>2</sup>

## 5 Discussion of the results

The most surprising finding in this study was that there is a tendency for better assessment scores in the Balkan name condition if there are no insertional switches into German. As soon as there are such inserts, the Balkan name texts are assessed in a consistently more negative manner, compared not only to the –inserts condition of the same name condition (see later), but also compared to the Swiss name +inserts condition.

Hypothesis 1 can thus be maintained with important reservations regarding the interaction of the two factors (hypothesis 3): Contrary to the initial hypothesis, the Balkan label itself does not consistently trigger more negative assessments of the texts. In the condition without code-mixing, the Balkan label actually leads to more positive evaluative judgments than the Swiss name label.

Given the tendency of exaggerating both in the positive and negative direction the assessments of the Balkan pupil's production, it is rather surprising to nevertheless find a main effect of the name on the assessments, since the more positive and more negative assessments could be expected to cancel it out. As we have already seen, this main effect is due to the more positive assessment of the cross-linguistic interferences dimension and to the more negative assessment of the academic potential dimension in the Balkan name label condition. Overall, at least in our data, we can observe a more sensitive reaction to the Balkan label, both in the +inserts and –inserts condition. Compared to the Swiss name conditions, in which the means of all five variables lie close to the middle ground, the Balkan label leads to both more positive and more negative 'exaggerated' assessments. In the light of the studies discussed at the beginning of this contribution it is certainly not a surprise that name stereotype effects were found in our data, but it is still important to note that in an ideal educational world with ideal raters the identity of the author of a text should not influence the assessment at all.

I assume that the Balkan label inevitably triggers stereotypical attributes as has been shown to be the case for other contexts (see e.g., Devine, 1989 for the inevitability of stereotypical first reactions): poverty, low education, therefore also limited school language proficiency (and hence also low proficiency in foreign languages), dishonesty, laziness and aggressiveness are the relevant stereotypes in our context (Stolz, 2000: 194). In order to explain the diverging results in the +inserts and –inserts conditions we can hypothesize that the participants' measurably positive reaction to the insertionally unmarked oral production in the Balkan condition is an expression of positive deception: stereotypically, one expects a migrant boy with a Serbian background to be overstrained by the foreign language education in school. Confronted with the actual oral production, at least in the –inserts condition, the proficiency level seems to be perceived as being 'surprisingly good'. Maybe the production is perceived as being – unexpectedly – free of 'Balkan-style' interferences, whatever they might be. This leads to the more positive assessment, which could thus be nothing else than a positive variant of a stereotypical reaction to the Balkan label. However, as soon as the insertional mixing is added to the stimuli, we can observe evaluative behavior in the form of strongly negative assessments of the texts, supposedly again triggered by the stereotypical point of departure. The inserts could be interpreted as 'proof' that the pupil with Balkan origin is overwhelmed by the challenge of learning the local language and foreign languages at more or less the same time. The indigenous label, on the other hand, does not activate the same stereotypical cognitive reference point, or if it did, then that point would be higher up on the stereotypical

competence scale. There is thus no positive exaggeration of the assessment in the – inserts condition, but there is also no exaggerated punishment in the +inserts condition.

The second main point concerns the perception and assessment of code-mixing phenomena. There is no doubt that the inserts have been noticed and dispreferred by the raters, since four out of the five variables in all labeling conditions show relatively speaking lower scores when inserts are present (holding the name factor constant). The raters are undoubtedly familiar with all kinds of surface manifestations of multilingualism in oral interaction since they are all students in a bilingual institution of a bilingual city in a quadrilingual country. Nevertheless, insertional mixing is devalued in assessment. As discussed earlier, there is a movement, at least in some western European contexts, towards more consideration for the nature and the dynamics of bi- and multilingualism in the (foreign) language classrooms. Especially in countries and regions with high proportions of migrants, which is the case in all urban centers in Switzerland, there is increasing emphasis on intercultural education and other implications of heterogeneous societies. The curriculum of the future primary school teachers who were the participants in this study clearly reflects this sensitivity.

The general influence of the four experimental conditions is similar for all five dependent variables analyzed. This can be interpreted as a Halo effect (Nisbett & Wilson, 1977): although not all the variables are directly and intrinsically related to insertional code-mixing, all variables react to changes in this independent variable. There is, for instance, no intrinsic causality between the word-finding problems in a foreign language that manifests itself in insertional mixing on the one hand and the global academic potential of the speaker on the other. On the contrary, given the change in pedagogical perspective on the monolingual paradigm in the language classrooms (see earlier discussion) one could even expect such inserts to be a manifestation of a communicatively sensible way of dealing with the limited repertoire in the target language, thus a sign of being a communicatively adequate and strategically versatile multilingual.

Such an interpretation could hypothetically apply to the Tim/Luca +inserts condition that – descriptively – shows even higher mean values for academic potential than the Tim/Luca -inserts condition (see Figure 3b). On the other hand, the tendency is very different in the Goran/Dragan condition, making the academic potential variable one of the two that show a main effect of the name factor. Only a very small number of students ( $N = 2$ ) refused to fill in the corresponding Likert item, which suggests that the item must have made sense to many of the participants, although they did not have other information than the ethnic label and the aural stimuli themselves. To a certain extent it is thus normal that the decontextualized experimental nature of the task creates such effects. But the important point here is that the negative impact of code-mixing on the academic potential only applies to the Balkan names, as if the future teachers do not expect

the Balkan boy to catch up (e.g., to learn sufficient French not to be forced to mix it with German), whereas the mixing does not seem to affect the ‘academic forecast’ of the Swiss boy. We can thus hypothesize that, as in the classical Halo effect research, features of speakers that have no intrinsic relation to the assessed dimensions ‘irradiate’ outside of their scope of application.

There are a number of issues that cannot be addressed based on the data gathered for this research. First, future research should investigate whether the effects would be different with other participants with more or less knowledge about bi- and multilingual development. Since we do not have analogous data from controls who are not studying in an environment that explicitly addresses issues of cultural and linguistic diversity, we cannot assess the impact (positive or negative) this education has on the outcome of the experiment. Obviously, it would also be important to see whether such controls or participants who have completed their studies or who have several years of professional experience as teachers would respond differently in a similar task. These would be relevant questions for future research.

Second, other groups of multilinguals in the labels, such as groups that are less stigmatized in German-speaking Switzerland, should be added (Anglophones, Francophones, and so on). This will allow determination of whether the mixing depreciation effect occurs with these labels as well, which we could indeed conclude is related to a general (stereotypical) folk theory of multilingualism, or if not, would mean that it is related to a particular, stereotypically negatively perceived group of migrants.

A last point that certainly is problematic in the research design chosen here is the limited ecological validity. Although the method of a field experiment allows at least the participants’ environment to be kept relatively natural, there is no doubt that actual assessment practices in the real school world are much more tightly connected to curricula, to preceding and following classroom activities, and do not occur in an isolated assessment event similar to this experiment. However, based on what we know from research on educational inequalities and the effects of teachers’ expectations (Jussim & Harber, 2005; Kronig, 2007 for the Swiss context), one should probably argue that the more the assessment is based on relevant data only, and the less contextual information the teacher has at her/his disposal, the less danger of expectancy effects there is. This obviously presupposes new ways of assessing language skills, and teachers should not only find ways of evaluating pupils’ texts anonymously, but above all should also have at their disposal a solid toolbox with workable tools for the assessment of their pupils’ productive and receptive language competences.

### Acknowledgements

I am very grateful to Alexandre Duchêne for his always eye-opening discussions of issues of social inequality and multilingualism, and I wish to thank Alexandre and his collaborators Sheila Pellegrini and Vania Widmer for opening up their seminars at the HEPFR for the data collection. Thanks also to Pascal Gygax and T. Florian Jaeger for their invaluable advice on statistical matters, and to two anonymous reviewers for their helpful comments.

This study was realized based on data gathered in a research seminar of the MA program in multilingualism at the University of Fribourg. I wish to thank the group of enthusiastic students who helped to design the experiment and to collect the data: Christine Ammann-Tinguely, Mirjam Andexlinger Felouati, Barbla Etter, Simone Ganguillet, Barbara Grüter, Isabelle Kälin, Anja Klein, Melanie Kutz, Evelyne Peter, Sonja Rittierott, Adrian Schori.

### Note

1. Initially, the plan was to run a  $2 \times 2$  design within each subject, thus presenting four texts in a row. Pretesting revealed that this would have taken too much time and only three text assessments could be performed.
2. The multivariate GLM analysis (the `manova()` function in the stats package of R) does not allow to take into account the fact that each rater assessed three texts, which means that the individual assessments are not fully independent of each other. An alternative procedure accounting for this problem was also applied, using the `lme4` package in R. The `lmer()` function allows modeling the raters and the texts as random factors (random intercepts) while using the same fixed factors as in the procedure described here. Such a model can be fitted for each dependent variable separately. The results of this procedure confirm the results discussed above: For the dependent variable of interferences, e.g., both main effects and the interaction between them yield absolute t values greater than 2, which shows statistically significant influence of the fixed factors (mixing:  $t = -5.463$ , name:  $t = 4.255$ , mixing x name:  $t = -3.433$ , see Baayen, 2009: 248).

### References

- Adler, M. (1977). *collective and individual bilingualism. a sociolinguistic study.* Hamburg, Germany:
- Helmut Buske Verlag. Auer, P. (1999). *code-switching in conversation : Language, interaction and identity.* London, UK: Routledge. Auer, P., & di Luzio, A. (1988). *Variation and convergence. studies in social dialectology.* Berlin, Germany and New York, NY: de Gruyter.

Baayen, R. H. (2009). *analyzing linguistic data: a practical introduction to statistics using r* (4th print ed.). Cambridge: Cambridge University Press. Backus, A. (2004). Convergence as a mechanism of language change. *Bilingualism: Language and cognition*, 7, 179–181. Berthele, R. (2011). On abduction in receptive multilingualism. Evidence from cognate guessing tasks. *applied Linguistics review*, 2, 191–220.

Bialystok, E. (1987). Influences of bilingualism on metalinguistic development. *second Language research*, 3, 154–166.

Bono, M., & Melo-Pfeifer, S. (2010, online first). Language negotiation in multilingual learning environments. *International Journal of Bilingualism*, (print version 2011, 15, 291–309).

Booth, A., Leigh A., & Varganova, E. (2009). *does racial and ethnic discrimination vary across minority groups? evidence from three experiments*. Sidney, Australia: Institute for Cultural Diversity.

Butzkamm, W., & Caldwell, J. A. W. (2009). *the bilingual reform: a paradigm shift in foreign language teaching* (Narr Studienbücher). Tübingen, Germany: G. Narr.

Byram, M., & Hu, A. (2009). Interkulturelle kompetenz und fremdsprachliches lernen: Modelle, empirie, evaluation [Intercultural competence and foreign language learning: models, empiricism, assessment]. Tübingen, Germany: G. Narr.

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism*, 7, 71–87.

Chertkow, H., Whitehead, V., Phillips, N., Wolfson, C., Atherton, J., & Bergman, H. (2010). Multilingualism (but not always bilingualism) delays the onset of Alzheimer disease: Evidence from a bilingual community. *alzheimer disease & associated disorders*, 24, 118–125.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of personality and social psychology*, 56, 5–18.

Diaz, R. M., & Klinger, C. (1991). Towards an explanatory model of the interaction between bilingualism and cognitive development. In E. Bialystok (Ed.), *Language processing in bilingual children*, (pp. 167–192). Cambridge, UK: Cambridge University Press.

Faerch, C., & Kasper, G. (1983). *strategies in interlanguage communication*. London, UK: Longman.

Gardner-Chloros, P. (2009). *code-switching*. Cambridge, UK: Cambridge University Press.

Greene, D. M., & Walker, F. R. (2004). Recommendations to public speaking instructors for the negotiation of code-switching practices among Black English-speaking African American students. *the Journal of Negro education*, 73, 435–442.

Grosjean, F. (1985). The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual and Multicultural development*, 6, 467–477.

Herdina, P., & Jessner, U. (2002). A dynamic model of multilingualism. perspectives of change in psycholinguistics. Clevedon, UK: Multilingual Matters.

Hughes, C. E., Shaunessy, E. S., & Brice, A. R. (2006). Code switching among bilingual and limited English proficient students: Possible indicators of giftedness. *Journal for the education of the gifted*, 30, 7–28.

Jessner, U. (2000). Metalinguistisches denken beim Drittsprachgebrauch – Bilingualismus ist kein zweifacher monolingualismus. In A. James (Ed.), *aktuelle themen im Zweitspracherwerb. Österreichische Beiträge* (pp. 77–88). Wien, Austria: Edition Praesens.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *personality and social psychology review*, 9, 131–155.

Kronig, W. (2007). die systematische Zufälligkeit des Bildungserfolgs. theoretische erklärungen und empirische untersuchungen zur Lernentwicklung und zur Leistungsbewertung in unterschiedlichen schulklassen. Bern, Switzerland, Stuttgart, Germany and Wien, Austria: Haupt.

Lenz, P. (2007). HarmoS FS – Bildungsstandards für den Fremdsprachenunterricht. *Babylonia*, 15, 24–29.

Lenz, P., & Berthele, R. (2010). assessment in plurilingual and intercultural education. satellite study No. 2 for the guide for the development and Implementation of curricula for plurilingual and Intercultural education. Strasbourg, France: Language Policy Division Council of Europe.

Li Wei (2004). *the bilingualism reader*. London, UK: Routledge.

Muysken, P. C. (2000). *Bilingual speech: a typology of code-mixing*. Cambridge, UK: Cambridge University Press.

- Muysken, P. C. (2006). Two linguistic systems in contact: Grammar, phonology, and lexicon. In T. K. Bhatia & W. C. Ritchie (Eds.), *the handbook of Bilingualism* (pp. 147–168). Oxford, UK: Blackwell.
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and social psychology*, 18, 62–85.
- Nisbett, R. E., & Wilson, T. D. (1977). Halo effect – evidence for unconscious alteration of judgments. *Journal of personality and social psychology*, 35, 250–256.
- Pavlenko, A., & Jarvis, S. (2002). Bidirectional transfer. *applied Linguistics*, 23, 190–214.
- Poplack, S. (1980). Sometimes I start a sentence in English Y TERMINO EN ESPAÑOL: Toward a typology of code-switching. *Linguistics*, 18, 581–618.
- Preston, D. R. (2010). Variation in language regard. In E. Zeigler, P. Gilles, & J. Scharloth (Eds.), *Variatio delectat: empirische evidenzen und theoretische passungen sprachlicher Variation (für klaus J. Mattheier zum 65. geburtstag)*. (pp. 7–27). Frankfurt am Main and Berlin, Germany, Bruxelles, Belgium, New York, NY, Oxford, UK and Wien, Austria: Peter Lang.
- Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and social psychology*, 18, 10–30.
- Ramirez, A. G., & Milk, R. D. (1986). Notions of grammaticality among teachers of bilingual pupils. *tesoL Quarterly*, 20, 495–513.
- Rebzani, M. (2005). Emplois dits ethniques. Juge-t-on la valeur du personnel sur la base de préjugés ou de compétences professionnelles? *Les cahiers internationaux de psychologie sociale*, 66, 43–53.
- Rosenthal, R., & Jacobson, L. (1992). *pygmalion in the classroom: teacher expectation and pupils' intellectual development* (newly expanded ed.) New York, NY: Irvington Publishers.
- Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *research in higher education*, 33, 511–531.
- Stolz, J. (2000). *Soziologie der Fremdenfeindlichkeit : theoretische und empirische Analysen*. Frankfurt, Germany and New York, NY: Campus.

Trim, J. L. M., & North, B. (2003). Common European framework of reference for languages: Learning, teaching, assessment (4th ed.) Cambridge, UK: Cambridge University Press.

Wagner-Egger, P., Gygax, P., & Ribordy, F. (submitted). Racism in soccer? Perception of challenges of black and white soccer players. *perceptual and Motor skills*.

Weinreich, U. (1953). *Languages in contact. findings and problems*. The Hague, the Netherlands and Paris, France: Mouton.

Weisgerber, L. (1966). Vorteile und Gefahren der Zweisprachigkeit. *Wirkendes Wort*, 16, 273–289.

### **About the author**

Raphael Berthele is Professor of Multilingualism at the Department of Multilingualism at the bilingual University of Fribourg/Freiburg, Switzerland. He is also on the board of directors and a founding member of the Fribourg Institute of Multilingualism. His research interests cover various aspects of multilingual language using and learning, both from a cognitive and a social perspective. His most recent work focuses on the empirical investigation of receptive multilingualism.