Karsten Schmidtke-Bode and Gregor Kachel*

Exploring the motivational antecedents of Nepalese learners of L2 English

https://doi.org/10.1515/iral-2017-0037

Abstract: This paper is the first to examine the motivational disposition of Nepalese learners of L2 English. Based on an adapted version of the questionnaire in (Kormos, Judit & Kata Csizér. 2008. Age-related differences in motivation of learning English as a foreign language: Attitudes, selves, and motivated behavior. *Language Learning* 58. 327–355. Doi:10.1111/j.1467-9922.2008.00443.x.), we test the robustness and culture-specific applicability of well-known motivational antecedents to this learner population, and we investigate how the effects of these antecedents are mediated by the learners' gender, age and regional aspects of the educational setting. In doing so, we offer novel ways of analyzing the data: Firstly, we employ random forests and conditional inference trees for assessing the relative importance of motivational antecedents. Secondly, we complement the traditional 'scale-based approach', which focuses on holistic constructs like the 'Ideal L2 Self', with an 'item-based approach' that highlights more specific components of such scales. The results are interpreted with reference to the L2 Motivational Self System (Dörnyei, Zoltán. 2005. *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum) and to previous studies on other Asian populations of L2 learners.

Keywords: L2 motivation, L2 motivational self system, English, Nepal, random forests, conditional inference trees

1 Introduction

It is universally acknowledged that motivation is one of the key non-linguistic predictors of differential success rates in L2 acquisition, outranked perhaps "only [by] aptitude" (Gass and Selinker 2008: 426). Not surprisingly, then, motivation has been a lively research area in SLA for several decades, with the

***Corresponding author: Gregor Kachel,** Fachbereich Sozial- und Bildungswissenschaften, University of Applied Sciences Potsdam, Kiepenhauerallee 5, Potsdam, 14469, Germany, E-mail: gregor.kachel@gmail.com
Karsten Schmidtke-Bode, Department of English, Leipzig University, Nikolaistraße 6-10, Leipzig, 04109, Germany, E-mail: kschmidtkebode@gmail.com

last 10–15 years bearing witness to a particularly intense upsurge of interest in the motivational antecedents of L2 learners. At the turn of the twenty-first century, researchers noted that some of the concepts that had dominated motivational research were problematic in view of the increasingly globalized nature of much L2 learning. Most importantly, Gardner's (1985) venerable notion of 'integrativeness', i. e. the learner's desire to "come closer to the [L2] community" (Gardner 2001: 5), no longer applied to languages like English, whose use as a global lingua franca had essentially dissociated the language from its specific speech communities. Moreover, in the widespread contexts of foreign language learning without a surrounding community to integrate into, the original notion of integrativeness can often contribute relatively little to explaining the motivational disposition of the learners in the classroom.

At about the same time that these concerns were raised, researchers also began to take up important new impulses for the study of motivation from psychology, notably from the theory of possible selves (Markus and Nurius 1986) and self-discrepancy theory (Higgins 1987). What emerged from these new directions of thinking was not only a number of innovative individual studies on language-learning motivation (e. g. Ushioda 2001; Noels 2003), but also a new comprehensive approach to the issue, known as the 'L2 Motivational Self System' (Dörnyei 2005, 2009). At the heart of this model are processes of self-imagery and vision, and the motivational impact of the language-learning experience as such. The former basically relate to "an individual's ideas of what they *might* become, what they *would like to* become and what they are *afraid of* becoming" (Dörnyei et al. 2014: 20), and motivated behaviour aims precisely at reducing the discrepancy between these imagined 'possible selves' and one's current actual self in relation to L2 competence. In this connection, the future self-image as a competent L2 speaker is known as the 'Ideal L2 Self', whereas the so-called 'Ought-to L2 Self' concerns "the attributes that one believes one ought to possess to meet expectations and to avoid possible negative outcomes" (Dörnyei 2010a: 80). Conceptualized this way, the Ideal L2 Self essentially induces a motivational promotion focus, while the Ought-to L2 Self is guided by living up to demands and the fear of failure and thus operates on the basis of a prevention focus (Higgins 1998). The language-learning experience, finally, influences motivation through the degree of enjoyment associated with a language course, its instructor and other experiential aspects of the immediate learning environment.

The formulation of the L2 Motivational Self System has generated an unprecedented amount of motivation research (cf. Boo et al. 2015 for a recent comprehensive overview). Underlying much of this research is the desire to test the applicability of the model to culturally diverse learning populations and to investigate the interaction of its central theoretical constructs with other variables in predicting motivated

learning behaviour. In keeping with these two objectives, the present study examines the motivational antecedents of an entirely novel population of L2 learners, viz. Nepalese learners of L2 English.[1] It is based on questionnaire data from roughly 150 learners from the Annapurna region. The questionnaire had originally been applied successfully to different age groups of Hungarian learners of English (Kormos and Csizér 2008) and was modified slightly to incorporate culture-specific aspects of the present learner population (cf. Section 2). While our study does not specifically target different age groups, it does take age-, gender- and region-based differences into account and thus contributes to recent discussions of how such factors interact with the different psychological constructs under scrutiny.

In order to ensure comparability with Kormos and Csizér's (2008) study, we follow a similar methodological approach: First, we apply factor-analytical techniques to probe the consistency of the theoretical constructs that the items on the questionnaire intended to capture (Section 3). For consistent and reliable scales, we then examine their relative influence on measures of motivated learning behaviour, using correlation and regression procedures (Sections 4 and 5). However, we here propose novel ways of analyzing the data in this way: On the one hand, we use alternative methods of regression by means of random forests and conditional inference trees, which allow us to estimate the relative importance of a large number of variables for a comparatively small number of observations. On the other hand, we complement the traditional 'scale-based' approach, in which summarized scales of questionnaire items act as predictors, by an 'item-based' approach that brings out (rather than conceals) how individual items on the questionnaire correlate with motivated learning behaviour. Throughout the paper, our findings are interpreted in light of similar studies on other Asian countries.

2 Data collection

2.1 The learner population

For the data collection process, the second author of this paper visited schools and colleges in the Annapurna region or contacted them via email.

1 To the best of our knowledge, there is no previous research on the motivational disposition of Nepalese language learners: In the recent thorough meta-analysis of motivational research between 2005 and 2014, Boo et al.'s (2015) list of 416 publications does not feature any entry on language learning in Nepal. The same holds for an even more recent volume dedicated specifically to L2 learners in Asian countries (Apple et al. 2016).

Out of 35 envisaged institutions, 9 granted permission to forward students a link to an online questionnaire ($n = 109$) or to visit students in the classroom and provide them with a paper-based version thereof ($n = 40$).[2] The majority of participants were recruited in Pokhara and will be referred to as the *urban* population ($n = 105$). With over 300,000 residents, Pokhara is the second largest city in Nepal (cf. also Figure 1 below). Serving as a starting point for hikes in the Annapurna region and being home to many schools and colleges, Pokhara is an important cultural and educational centre in the area and a very popular destination for tourists. In Pokhara, we obtained data from

- 59 secondary school pupils (age range = 12–17 ($\bar{x} = 14.89$, $sd = 1.54$), 27 female) attending five different secondary or boarding schools[3]
- 46 university students (age range = 18–39 ($\bar{x} = 22.69$, $sd = 3.67$), 18 female[4]). The university students attended classes in chartered accounting at the Enlightened International Academy Pokhara ($n = 13$) or were students of

Figure 1: Geographic location of Pokhara and Nangi (from HEF website[5]).

2 These and all subsequent figures relate to the final number of questionnaires that was taken into account. Some forms had to be excluded due to more than five missing values or for other reasons.

3 Schools ($n = 5$): Jyoit Vocational Training Center, National Inventive Boarding School, Pokhara Creative English School, Shree Himanchal Higher Secondary School, Shree Krishna Lower Secondary School.

4 Five of the university students did not to provide information on their gender.

5 cf. http://himanchal.org/nangi-village (accessed 2 May 2016).

English ($n = 26$) and other fields ($n = 7$) at the Prithvi Narayan Campus Pokhara, an educational centre of Tribhuvan University, one of the largest public universities of the country.

These two groups can be contrasted with the *remote* subset of participants: It comprises 44 secondary school pupils (age range = 16–20 ($\bar{x} = 18.34$, $sd = 1.06$), 27 female) attending Himanchal Higher Secondary School (HHSS) in Nangi village (cf. Figure 1 below). HHSS is a boarding school hosting over 300 students from several settlements in the Annapurna region, most of which are entirely without access to basic infrastructure. Agriculture, tourism and crafts are the main sources of income in the region. Owing to the work of the Himanchal Education Foundation (HEF), Nangi village has grown into a prototype for community-based development in Nepal. HEF is devoted to creating income-generating projects like yak-farming or handmade paper production, the proceeds of which support the local boarding school. HEF has installed a Wi-Fi relay station providing students and teachers with Internet access in the local solar-powered computer lab, although the power supply for the village only lasts for a few hours each day. HEF also invites volunteers from all over the world to contribute with their projects. The participants of this group are a unique population as they come from a remote area and live under very simple conditions but at the same time enjoy formal education and have access to social media and English cultural products via the Internet.

2.2 The survey instrument

The present study is based on psychometric data, specifically the questionnaire employed by Kormos and Csizér (2008) for Hungarian learners of English: It intends to capture the relative impact of a large number of antecedents (i. e. so-called motivational scales like "parental encouragement") on the participants' degrees of motivated learning behaviour. To this end, each scale is operationalized by several distinct items on the questionnaire. For the most part, these are formulated as statements that the participants were asked to (dis)agree with on a five-point Likert scale (e. g. "I would get tense if a foreigner asked me for the way in English"). Overall, the original questionnaire comprised 69 items mapping onto 14 different motivational scales.[6]

6 We would like to thank Judit Kormos and Kata Csizér for providing us with their original questionnaire, as well as for making part of their SPSS code accessible. We are also grateful to Stephen Ryan for sharing his questionnaire (Ryan 2008) for comparison.

For the present investigation, we used a slightly modified version of the original questionnaire: To begin with, the questionnaire was expanded by several items in order to incorporate two further scales. One of these is the construct of self-efficacy, which captures the learners' perceived degree of control over the "attainment or avoidance of a possible self" (Dörnyei 2009: 19). After all, an ideal self concept need not only be attractive but also realistic in order to be an effective motivator (cf. Ruvolo and Markus 1992: 96 on this point). The second scale we added pertains to the perceived difficulty of the target language: No matter how beneficial or prestigious it may be to become a competent speaker of English, a student might easily be demotivated by the learning effort she would have to invest in this goal if the L2 is perceived as particularly challenging to master. A further way in which the original questionnaire was altered relates to either smaller modifications of individual items (e. g. removing them if they are not applicable to the learning context at hand) or to the allocation of items to scales; these will be discussed below.

In the remainder of this section, we now briefly introduce the various scales. In each case, we outline how many and which items belong to the scale in question, whether the application of the scale differs from Kormos and Csizér (2008) and where other changes to the survey instrument were made in order to adapt it to the present population. For more detailed information, the reader is kindly referred to Appendix 1, which features a sorted version of the full questionnaire.

- **Integrativeness** (items 1–3): Degree of identification with the L2 and its speakers, mostly in Gardner's original sense (e. g. "How much would you like to become similar to the people who speak English?").
- **Attitudes towards the L2 speakers and the community** (items 4–8): Degree of affection for the L2 culture and the desire to (travel to) meet native speakers.
- **Instrumentality** (items 9–12): Anticipated personal advantages from learning English, such as the helpfulness of English for future travelling and job opportunities.
- **Cultural interest** (items 13–15): Attitude towards L2 cultural products (here: movies, magazines and pop music made in the USA, following Kormos and Csizér (2008); a question relating specifically to television was removed because of the learners' very limited access to TV in general).
- **Vitality of the L2 community** (items 16–19): Students' estimation of the global importance, wealth and developmental status of the UK and the USA.

- **Linguistic self-confidence** (items 20–22): Students' views on how easily and successfully they will be able to acquire a foreign language.
- **Language use anxiety** (items 23–26): Students' estimation of the stress they would feel if they had to speak English in a variety of situations in everyday life.
- **Classroom anxiety** (items 27–30): Students' estimation of their anxiety in English lessons.
- **Milieu** (items 31–34): the attitude of people in the students' environment concerning the importance of learning English and foreign languages more generally.
- **Parental encouragement** (items 35–38): The degree to which parents encourage their children to learn English.
- **Language-learning attitudes** (items 39–42): The degree of enjoyment or satisfaction that students derive from the language-learning process (e. g. "I really enjoy learning English."), but also their attitudes to learning foreign languages more generally ("I think that foreign languages are important school subjects.").[7]
- **International posture** (items 43–47): The degree of openness towards using English to communicate with people from all over the world, not just English-speaking countries.
- **Ideal L2 Self** (items 48–54): Students' future views of themselves as successful L2 speakers, as outlined in the introduction to this paper (e. g. "I like to think of myself as someone who will be able to speak English."). Note, however, that the original questionnaire, as well as similar ones that have been used in the literature (e. g. Ryan 2008 on Japanese learners), also include items like "The things that I want to do in the future *require* me to speak English." (here: items 55 and 56). We feel that these items are rather different in nature and we shall return to them in later sections. For the sake of research comparability, however, we left them in the Ideal L2 Self category.
- **Ought-to L2 Self** (items 55–59): Across the literature on motivation, the most typical conceptualization of this scale is that it reflects "motives generated by a sense of duty or a fear of punishment", so as to avoid negative outcomes (Csizér and Dörnyei 2005: 29), although other interpretations can also

7 The name of the scale is adopted from Gardner (1985) and Kormos and Csizér (2008). However, most of its items effectively cover Dörnyei's (2005, 2009) notion of the 'language learning experience', which (apart from the Ideal and Ought-to L2 Self) is the third central cornerstone of the L2 Motivational Self System.

be found.[8] Accordingly, Kormos and Csizér (2008) include the fear of letting other people down (item 55), the perceived necessity of learning English because it is an international language (57) and because it is required to be considered an educated person (58). In addition, they include the fear that learning English will make one feel less 'native' with regard to one's L1 community (56), and the assessment of whether anybody cares whether one learns English or not. This last question, however, does not have the same deontic force as the other ones, which is why we felt it more appropriately placed in the category 'milieu' (32). Instead, what we added to the Ought-to L2 Self is item 59, which is concerned with the fear of disappointing one's teacher. This more specific instance of 'letting other people down' may be highly relevant in a cultural setting where teachers enjoy particular prestige and where students' efforts may thus be directed at duteous learning behaviour and at maintaining a respectful relationship with their teacher (cf. also Gobel et al. 2016 on this point).

- **Self-efficacy** (items 60–67): The students' perceived degree of control of and responsibility for their performance. This scale was added to the original questionnaire because research in psychology suggests that self-efficacy is central in defining learning goals and thus constitutes a precondition for motivated learning behaviour (cf., e. g., Zimmerman et al. 1992; Zimmerman 1995; Bong and Clark 1999; Kormos et al. 2011). Following a suggestion by Jonkisz et al. (2011), our scale of self-efficacy consists of complementary positively and negatively coded items (e. g. "I am responsible for the quality of my English" and "I am not responsible for the quality of my English").[9]
- **Perceived L2 difficulty** (items 68–71): Another new scale that measures the learners' perception of the difficulty of English as an L2.
- **Motivated learning behaviour** (items 72–82): The learners' efforts and persistence in learning English. This scale thus captures the degree of motivation and will hence serve as the dependent (or response) variable in the regression analyses. The original scale was enhanced by two further

8 For example, in Csizér and Kormos (2009: 103), a narrower view is adopted in which the Ought-to L2 Self refers to "students' perceptions of how important learning is in the opinion significant others", thus overlapping with what is called 'milieu' (and possibly 'parental encouragement') here. A general problem in the questionnaire-based literature on motivation is that the assignment of items to scales is not always made fully explicit, making it hard to replicate studies exactly.

9 Research in test theory has found that such complementary items never fully correlate and that especially with regard to measures of perceived control, participants tend to 'acquiesce', i. e. to respond positively throughout (regardless of the item's content), cf. Jonkisz et al. (2011: 61 f.).

items (81 and 82), which enquired the learners' will to engage with English outside the classroom and to study more than what is generally expected from them.

In addition to the above items, we collected information on the following biographical variables:

- **Age** (numerical)
- **Gender**: 72 female and 72 male subjects (5 subjects did not provide their gender)
- **School category**: 103 pupils versus 46 university/college students
- **Area**: 44 remote versus 105 urban learning environments
- **Number and kind of foreign languages learned**: For 106 participants, English is the only foreign language they learn, and English, Hindi and Japanese are the most commonly learned foreign languages in the sample.

The questionnaire was translated from English into Nepali by a native speaker of Nepali teaching English literature at university level. The Nepali version was then retranslated and checked for possible ambiguities by two native speakers teaching English at secondary schools.[10]

3 Identifying latent dimensions of variation

Following the analytical procedure in Kormos and Csizér (2008), the first question is whether the individual items that were meant to represent an underlying motivational scale do actually correlate with each other in the intended way and can thus be said to reflect the same latent dimension. A corollary of testing for this question is that it allows us to assess how well the original survey instrument carries over to a completely different learner population. In addressing these issues, we complement Kormos and Csizér's (2008) principal-component analysis (henceforth PCA) by exploratory factor analysis (FA); although these two techniques of dimension reduction are closely related and often yield

10 The participants were provided with the questions in Latin script but this did not hinder comprehension as all participants were highly familiar with reading and writing Nepali in Latin script on the internet and various social media. The online version of the questionnaire was created and hosted by drawing on the resources of the Department of Educational Sciences at Friedrich Schiller University Jena, the university at which both authors worked at the time of data collection. We would like to thank this department, as well as the Department of English and American Studies in Jena, for their support.

similar results, FA "is more appropriate for detecting theoretically relevant underlying dimensions in the data" (Levshina 2015: 352).[11]

Dimension-reduction procedures are complex, multi-step statistical methods. In order to keep the presentation concise and readable, we outsource a detailed description of our modus operandi to Appendix 2, while the present section focusses on the results of the analyses. As a rough orientation, the two dimension-reduction techniques basically scan the correlational structure of the items pertaining to a given motivational scale and ideally return that they can be reduced to a single 'principal component' or 'factor'. Sometimes, such homogeneous solutions can only be obtained if individual 'distorting' items are removed from a scale, and it is also possible that a purported scale actually falls apart into several distinct factors (cf. also Appendix 2). Finally, and in keeping with Kormos and Csizér (2008), the scales obtained were checked for their reliability, using Cronbach's alpha as a criterion measure.[12] In Table 1, we present the results of this procedure: The table is divided into reliable, homogeneous scales on the left and diverse ones on the right, i. e. those that did not yield a convincing one-factor solution and/or sufficient reliability. Moreover, the typography of the individual scales in the table indicates whether the results are in accordance with or deviate from the ones obtained by Kormos and Csizér (2008).

As can be seen, when compared to Kormos and Csizér (2008), there is much overlap in the results for the homogeneous scales, despite the fact that a few individual items may have been allocated differently in the first place or treated differently during the factor analysis. In contrast to the original study, we also obtained a fairly reliable scale for instrumentality and for the vitality of the L2 community. Interestingly, the latter had disintegrated into a UK-related scale and a USA-related scale in Kormos and Csizér (2008), while our subjects showed a much more unified perception of the importance of the two countries in the world today. The table also shows that our new scale of the perceived L2 difficulty emerged as a reliable gauge.

As for the heterogeneous scales, our results concur with Kormos and Csizér's (2008) only as far as linguistic self-confidence and the Ought-to L2 self are

11 Based on Stevens (2002), Field et al. (2012: 759) suggest comparing the results of both techniques if the number of dimensions to be reduced is rather small, as in the present case, so we will follow this complementary procedure here. Field et al. (ibid.) also offer a more general discussion of the similarities and differences between FA and PCA.

12 As explained in Appendix 2, we follow Kormos and Csizér (2008) and other studies on motivation in taking $\alpha \geq 0.7$ as criterial; however, slightly lower values (e. g. 0.68) are still acceptable if the number of items is small (cf. also Dörnyei 2010b: 95).

Table 1: Results of the latent-dimension and reliability analyses.

Homogeneous and reliable scales			Heterogeneous and/or unreliable scales		
Scale	**Items**	**α**	**Scale**	**Items**	**α**
Attitudes to the L2 community	4–7	0.71	Integrativeness	1–3	0.48
Instrumentality	9–11	0.68	Cultural interest	13–15	0.57
Vitality L2 community	16–19	0.78	**Ling. self-confidence**	20–22	0.19
Language use anxiety	23–26	0.74	Milieu	31–34	0.09
Classroom anxiety	27–29	0.73	Learning attitude/ experience	39–42	0.51
Parental encouragement	35,37,38	0.74	International posture	43,44,46,47	0.55
Ideal-L2-Self	48–54	0.71	**Ought-to-Self**	55–59	−0.52
Perceived L2 difficulty	68–71	0.7	*Self-efficacy*	60–67	0.06

Legend: regular typeface = homogeneous scales in Kormos and Csizér (2008), **bold** = heterogeneous scales in Kormos and Csizér (2008), *italics* = not part of Kormos and Csizér's (2008) study.

concerned. Both of these constructs need to be captured by at least two factors each, pointing to a heterogeneous statistical structure of the underlying items. Just as in the original study, then, one of the three cornerstones of the L2 Motivational Self System does not receive empirical support in our population (but cf. Section 5 for discussion). Interestingly, however, this also holds for the purported scales of integrativeness, cultural interest, milieu, learning attitudes, self-efficacy and international posture, none of which achieves the required degree of unity and/or reliability (see Appendix 2 for details). For our newly added scale of self-efficacy, both PCA and FA suggest a very complex factor structure (with at least three factors for the eight items in question), and even the most homogeneous of these factors (items 60, 64 and 67) fails to pass the reliability test ($\alpha = 0.56$). A similarly unruly picture holds for the items coding milieu, whose statistical structure does not even fulfil the basic requirements for factor-analytical procedures, regardless of whether the worst-fitting item (32) is removed. Perhaps surprisingly, the three-item scale probing various cultural products of the USA is also somewhat problematic: While a PCA (but not a FA) warrants a one-component solution, this scale, too, fails the reliability test ($\alpha = 0.57$). Essentially the same applies to international posture and integrativeness. With regard to the latter, our results are thus more in line with Taguchi et al.'s (2009) for other Asian populations than with Kormos and Csizér's (2008) for Hungarian learners: The alpha values for integrativeness are the lowest in Taguchi et al.'s study (ranging between 0.56 and 0.64 for the different subpopulations), and ours is even lower (0.48). Finally, the intended scale capturing

language learning attitudes and experience disintegrates in unexpected ways: The conceptually very similar items 39 and 40 do not correlate very highly, and an acceptable factor solution and alpha value can only be achieved if the scale is reduced two items (39, 42, $\alpha = 0.7$), which actually constitute a mix of attitudes and experience. This is in stark contrast to Kormos and Csizér (2008), whose final calculation included the purely experience-related items 39–41 and yielded high alpha values in all of their subsamples.

Overall, then, our results suggest that some of the antecedent scales of Kormos and Csizér (2008) can also adequately describe our sample of Nepalese learners of English, although some crucial differences emerge as well. With regard to the cornerstones of the L2 Motivational Self System, the findings align in terms of the Ideal L2 Self (strongly homogeneous) and the Ought-to L2 self (strongly heterogeneous), but do not fully accord on the language-learning experience (reliable only if attitudes to foreign languages are included, as in Gardner's original conception of the construct).

To round off the present section, we should finally also look at the scale of motivated learning behaviour as such, i. e. the dependent variable of the subsequent sections. Here it turns out that our addition of two items to the original scale (81 and 82) introduces heterogeneity; removing these items renders a one-factor solution more acceptable (though still not ideal, cf. Appendix 2) and also increases the alpha value to a satisfactory 0.83. In all following analyses, we will hence only take the original items 72–80 into account.

4 Group-specific statistical structure of the scales

We are now going to examine the statistical structure of the reliable scales in some more detail. Specifically, we will outline how the various subgroups in our data (i. e. the different regions, genders and age groups) differ with regard to the mean scores on the motivational scales and also with regard to how each antecedent scale correlates with the dependent variable of motivated learning behaviour (i. e. the mean of items 72–80), thus paving the way for the regression analyses to come. The pertinent data are summarized in Appendix 2.

To begin with, the degree of motivated learning behaviour of our participants can be characterized as extremely high ($\bar{x} = 4.22$, $sd = 0.54$), as compared to both the Hungarian learners of Kormos and Csizér (2008) ($\bar{x} = 3.73$) and other Asian learner populations (e. g. Japan ($\bar{x} = 3.08$ in Taguchi et al. 2009), China ($\bar{x} = 3.63$ in Taguchi et al. 2009; $\bar{x} = 3.6$ in You and Dörnyei 2016) and Iran

($\bar{x} = 3.79$ in Taguchi et al. 2009).[13] Perhaps the key difference to these countries is that English plays a crucial role in the education and everyday lives of our subjects: It is a prerequisite for enjoying higher education at the urban universities, and even the remote subsample of our participants regularly encounters English speakers (though not necessarily native speakers) outside the classroom, owing to a teacher exchange and volunteer programme as well as tourist facilities in Nangi village. In contrast to other studies, then, the target language is much more present in the participants' lives.[14] As far as the different subsamples are concerned, the only significant difference in motivated learning behaviour was found between pupils and students, with the latter showing an even higher degree of learning effort than the former (see Appendix 2 for the data). Considering that a large portion of these students is enrolled in an English degree programme, this result is probably not surprising, but it should be noted that the same trend has also been observed in completely different samples (e. g. in Kormos and Csizér 2008).

Let us now turn to differences between the various subsamples on the *antecedent* scales. Starting with gender differences, it turns out that most of our measurements are rather homogeneous across the two gender categories.[15] Notable exceptions are found in the significantly higher degree of parental encouragement reported by the female subjects, and by certain correlations of antecedents with motivated learning behaviour. Specifically, favourable attitudes to the L2 community correlate with learning efforts only in the female subsample, while the male group does not yield a significant statistical signal here. Conversely, language-learning attitudes and experiences correlate with motivated behaviour much more strongly in the male group (although the correlation is significant for both males and females). Finally, the correlation between perceived L2 difficulty and motivation is only (marginally) significant

13 Note that the studies on Asian populations used 6-point Likert scales, which makes a direct comparison of mean values difficult. The means reported here are downscaled to fit a 5-point scale and hence need to be taken with a grain of salt.

14 By contrast, it is unlikely that this high average value of the motivational scale is due to unreliable scoring since the scale comprises not only one but nine items, and the previous section showed that these are coherent in their correlational structure.

15 This is in stark contrast, for example, to You and Dörnyei's (2016) findings in China, where nearly all scales examined show significantly higher average scores for females. Note, however, that the two studies are not directly comparable, as the scales (and the internal composition of the scales) are partly different and – perhaps most importantly – the sample size for China is much larger than in the present case. Therefore, it may be that certain well-researched gender biases in L2 motivation (cf., e. g., Henry 2011; Henry and Cliffordson 2013) simply fail to show in our smaller sample.

for the female group. Apparently, then, our female subjects are relatively more likely to be discouraged in their learning endeavours if the L2 presents itself to them as a difficult object of study.

Contrasting the remote and urban parts of the sample, the significant differences can be summarized as follows: Subjects from the remote area show higher mean scores for instrumental motives and parental encouragement, but lower mean scores for the language-learning experience (e. g. for their joy in learning English) and for the vitality they attribute to the USA and the UK. As far as correlations with motivated behaviour are concerned, we found that only the subjects from the remote subsample show a significant negative correlation with both language-use anxiety and classroom anxiety. Thus while the mean level of both anxiety types is very similar in the remote and urban populations, it is only in the former that this actually correlates with the subjects' learning motivation. Additionally, it turns out that the correlation between the perceived L2 difficulty and (de)motivation is much stronger for the remote part of the sample (it only borders on significance for the urban sample), and that there is an even stronger association between the Ideal L2 self and motivated behaviour in the remote sample (although it is highly significant in both). Overall, then, the regional differences that we find in our data are quite varied and rather subtle in their effects, so that they do not primarily reflect a cultural difference in how "integrated into the globalized international world" the two areas are (You and Dörnyei 2016: 508).[16]

Finally, let us turn to the contrast between pupils and students; recall that this is not coextensive with the remote-urban distinction, as the urban subsample also contains a large number of secondary-school pupils. We saw above that the remote subsample exhibits a higher average degree of instrumentality; it is thus interesting that this is now reversed in terms of age: While instrumental motives are very strong in both samples, they are slightly lower for the pupils. Additionally, there is also a difference with regard to the Ideal L2 Self: While achieving high average scores in both subsamples, the construct yields a significantly stronger signal for the students; in this respect, our population is highly similar to the Hungarian learners studied by Kormos and Csizér (2008). As a final point, let us look at the correlations with motivated learning behaviour: There are only two scales for which this correlation is significantly different for pupils and

16 Recall from above that the remote area in our study provides internet access and (some) opportunities to interact with L2 speakers. This makes it arguably very different from a genuinely rural and underdeveloped area, even though there may still be significant disparities in the socioeconomic backgrounds of the urban and rural participants of our study (cf. also Kormos and Kiddle 2013 for further discussion of this issue).

students, viz. classroom anxiety and perceived L2 difficulty. In both cases, there is a significant negative correlation only in the younger subsample, while no significant signal can be obtained from the older learners. This can be interpreted in light of previous research, which has found that university students (as opposed to pupils) are generally less likely to be deterred by negative affect, which may in turn be due to a gain in self-confidence throughout adolescence (Carlson 1965), a greater level of perceived autonomy in university settings (Kormos et al. 2008) and, finally, better language abilities resulting from prolonged L2 education and more communicative experience (Yashima et al. 2004).[17]

5 Predicting motivated learning behaviour

What emerges from the correlational data in the previous section is a rather complex picture in which several scales co-vary with motivational behaviour, sometimes in interaction with gender, age and area effects. Clearly, then, what is needed is a multivariate statistical model that can shed light on the causal structure behind these different variables. In devising such a model, Kormos and Csizér (2008) ignored the scales that their PCA had identified as inconsistent/unreliable and, accordingly, they built a model in which the mean values of all reliable scales were taken as predictors for the mean of the motivational scale. In our first model to be presented below, we follow this procedure (but add gender, age, area and individual as predictors). However, operating only with the reliable scales and only with mean values for all predictors abstracts away from the actual data, and we have reason to believe that this procedure may miss some critical information: In particular, while scales such as the Ought-to L2 Self did not come out as a unified latent dimension, it is still the case that individual items on that scale correlate with several of our motivational items simultaneously. For example, students who believe that they should be able to speak English to be considered an educated person also tend to show motivated learning behaviour along several of the relevant items.

17 Apart from the correlations reported on in the main text, we also ran a few other, potentially interesting explorative analyses, with the following results: There was no significant overall correlation between motivation and the number of foreign languages the students were studying (or had studied) (Spearman's $rho = -0.116$, $p_{\text{one-tailed}} = 0.92$); the number of these foreign languages was not significantly different for the two areas (Wilcoxon test, $W = 2514$, $p_{\text{two-tailed}} = 0.396$). Similarly, we found no correlation between the number of these foreign languages and the degree of perceived L2 difficulty (Spearman's $rho = -0.133$, $p_{\text{two-tailed}} = 0.11$), so that, for example, knowledge of L2s other than English (e. g. Japanese or French) did not systematically affect the perception of the difficulty of English as an L2.

In order to probe such item-specific effects, we will run a second model in which individual items are taken directly as predictors of motivated learning behaviour. To this end, we first examined the correlational structure of all individual items in our questionnaire with the criterion items 72–80. Where any of those individual antecedent items showed correlations above *rho* ≥ 0.4 with *several* of the motivational items, we considered this to be a potentially interesting predictor variable. For example, item 11 from the instrumentality scale (inquiring the perceived importance of English in the world these days) shows such stronger correlations with three items on the motivational scale, while other instrumentality items (like 9 or 10) do not. Consequently, taking the mean of all instrumentality items as a predictor of motivated learning behaviour may seriously distort (i. e. underestimate) the influence of specific aspects of instrumentality. In the second model below, then, only individual items will thus be taken as predictors.

In predicting motivated learning behaviour, we here apply non-parametric regression methods known as random forests (henceforth RF) and conditional inference trees (CIT). These are more suitable to our data than traditional multiple regression, as they are "especially useful in the presence of many high-order interactions and in situations when the sample size is small, but the number of predictors is large" (Levshina 2015: 291, see also Tagliamonte and Baayen 2012; Wiechmann 2011; Lohmann 2013, for different linguistic applications of this method, and Strobl et al. 2009b, for a more general introduction to the underlying statistics). In a nutshell, CITs work on the basis of binary recursive partitioning of the data, splitting the dataset successively according to the predictor variables that show the strongest association with the response variable. In this way, one can obtain a tree-structured statistical model of the data which brings out all significant predictors and their relative importance. CITs do not make distributional assumptions and use permutation methods to determine the p-values given a predefined level of significance (here: $p < 0.05$). This, again, makes them attractive for our heavily non-normal data, our relatively small sample size given a substantial number of predictors, and possible outliers on each item or scale. A random forest (Breiman 2001) then consists of a large number of CITs (here: 3,000) that are grown on subsets of the data, i. e. by random sampling from cases and predictor variables. The specific implementation of RFs we use (Strobl et al. 2009a) is described in Appendix 2. In the following sections, CITs and RFs will be employed as complementary tools: While a single CIT on the whole data set nicely visualizes how the significant predictors successively split the sample into several groups, RFs afford a more stable estimate of the relative importance of all predictors in the model.

5.1 Scale-based regression analysis

As was mentioned above, the first regression model includes the mean value of the motivational items 72–80 as the response variable and the reliable scales from Table 1 (plus the two-item scale of learning attitudes/experience) as numerical predictors. In addition, we include gender, area, age category and individual[18] as binary predictors. If we fit a single CIT to our dataset, we obtain the tree structure displayed in Figure 2.

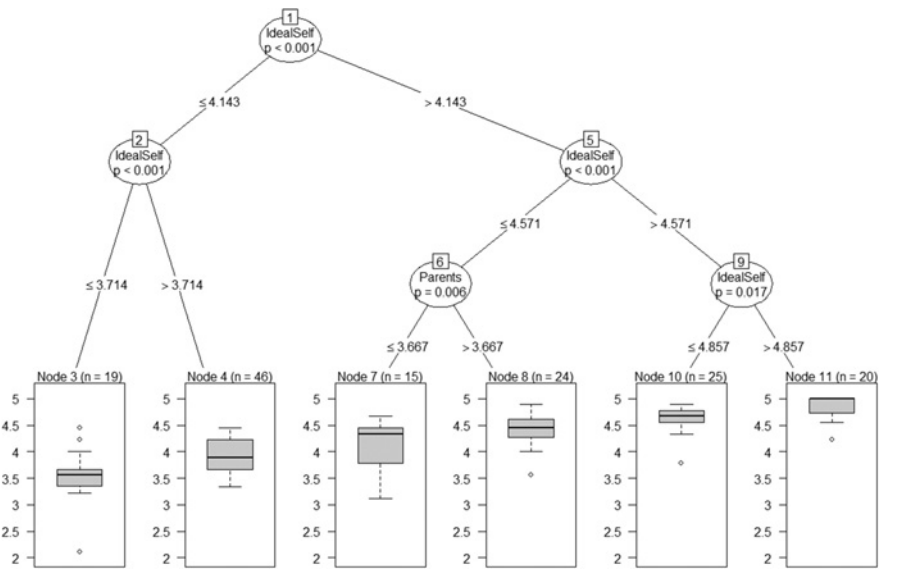

Figure 2: CIT for the scale-based model of the data ($R^2 = 0.617$).

As can be seen, the binary recursive algorithm has split our sample into six subgroups, represented by the bins at the bottom of the graph. The predictors that induce successive splits in the data are represented by the ovals at each node of the tree. In this model, there are only two significant predictors that

18 Every participant in our study is uniquely identified in our data by an ID. Since, by definition, every participant gave multiple ratings in the questionnaire, we are dealing with a within-participant design that requires us to include 'individual' (= ID) as a variable in the regression process. In mixed-effects modelling, this is normally done by treating individuals as random effects, and in the tree-based models to be presented here, we follow Tagliamonte and Baayen (2012) in simply including 'individual' into the equation for trees and forests, alongside all other factors.

partition our data: The strongest predictor is the scale of the Ideal L2 Self. It first splits the data into 65 cases on the left-hand side (the two leftmost bins), and 84 cases on the right (the four rightmost bins). The latter show that very high mean ratings on the Ideal L2 Self scale ($\bar{x} > 4.14$) induce the highest values on the motivational scale (see the boxplots in the rightmost bins). Within this group, it is again the Ideal L2 Self that defines two subgroups of highly motivated students, comprising 39 and 45 subjects, respectively. The latter subgroup is defined by an extremely high mean value of the Ideal L2 Self ($\bar{x} > 4.57$), and is itself split again in much the same way by the highest values on the Ideal L2 Self scale ($\bar{x} > 4.87$). Apparently, then, the highest degrees of motivation in the data are determined by this variable, thus confirming its importance in recent theories of L2 motivation.

The second scale that exerts a significant influence on the distribution of the response variable is parental encouragement: For the 39 subjects with a slightly lower mean Ideal L2 Self rating ($4.14 < \bar{x} \leq 4.57$), this factor induces a split such that higher levels of parental encouragement also lead to significantly higher degrees of motivated learning behaviour. In contrast to the Ideal L2 Self, this factor did not play a decisive role in Kormos and Csizér's (2008) study on Hungarian learners, and its significance for other Asian learner populations is hard to discern since neither Taguchi et al. (2009) nor You and Dörnyei (2016) investigated parental encouragement but mostly parental *expectations*, which are closely related to the Ought-to L2 Self.

Our CIT on the data can now be complemented by a random forest in order to arrive at a model that is less dependent on potential idiosyncrasies of the specific sample at hand. Because random forests work with many permuted subsets of the data (i. e. various learning and test sets containing only a certain portion of the observations and the predictors), they may yield a slightly different and often more complex picture of the variables that have an impact on the criterion measure. This is illustrated in Figure 3, which displays the so-called conditional variable importance for a random forest on our data (cf. Appendix 2 for more technical details).

Permuting the data in this way confirms the overwhelming importance of the Ideal L2 Self, whose influence outstrips that of all other predictors by far. However, Figure 3 suggests that at least four other factors return significant importance scores: In addition to the parental encouragement from above, we also see attitudes to the L2 community, instrumentality and attitudes to L2 learning emerging as significant antecedents. The influence of language-learning attitudes and experiences resonates with the results in Kormos and Csizér (2008), and it is also immediately plausible why a positive attitude towards speakers of English, towards travelling to English-speaking countries and the

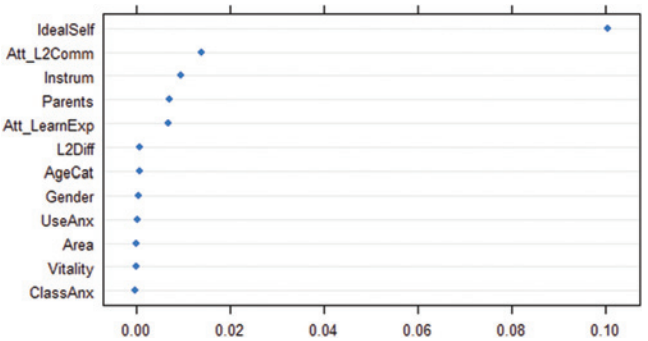

Figure 3: Conditional variable importance in the scale-based model of the data ($R^2 = 0.689$).

instrumental motives associated with a command of English would correlate with motivated learning behaviour. The random forest thus also supports some of the classic Gardnerian ingredients of L2 motivation, although it is indeed the case that the Ideal L2 Self acts as the strongest pull force in this complex mix.[19] Importantly, this immense influence remains stable over different random-forest models of the data, while the relative ranking of, for example, parental encouragement and instrumentality varies over several computations. Moreover, we will argue below that most of these additional factors are actually irrelevant if a more fine-grained, item-based model of the data is used, which is what we are turning to now.

5.2 Item-based regression analysis

In constructing an item-based model of our data, we followed the procedure outlined in Section 5.1 above. The relevant predictor items are listed in Table 2. As in the previous model, we added gender, area, age category and individual as binary predictors. The CIT obtained for this model is shown in Figure 4.

19 Following established criteria in the literature (cf. Appendix 2), some of the lower-ranking variables in Figure 3 (such as, for example, L2Diff and AgeCat) could also be said to be significant predictors, but since these are so close to the minimum threshold for significance in this model and clearly lag behind the 'truly' significant factors, we will neglect their influence here. This is further supported by the fact that their impact is not stable across several random forests: In some permutations, they simply do not come out as significant predictors.

Table 2: Predictors for the item-based model of the data.

Item	Scale	Description
X7	Att_L2Comm	How much would you like to travel to the USA?
X11	Instrumentality	How important do you think English is in the world these days?
X12	Instrumentality	Knowledge of English would make me a better educated person.
X40	Att_L2 Learning	I really enjoy learning English.
X41	Att_L2 Learning	I find learning English really interesting.
X44	IntPosture	If I could speak English well, I could get to know more people from other countries (not just English-speaking countries).
X45	IntPosture	I would like to be able to use English to communicate with people from other countries.
X48	Ideal L2 Self	I like to think of myself as someone who will be able to speak English.
X49	Ideal L2 Self	If my dreams come true, I will use English effectively in the future.
X50	Ideal L2 Self	Whenever I think of my future career, I imagine myself being able to use English.
X51	Ideal L2 Self	When I think about my future, it is important that I use English.
X52	Ideal L2 Self	I can imagine speaking English with international friends.
X53	Ideal L2 Self	The things I want to do in the future require me to speak English.
X54	Ideal L2 Self	The job I imagine having in the future requires that I speak English well.
X58	Ought-to L2 Self	For me to be an educated person I should be able to speak English.
X59	Ought-to L2 Self	I study hard in order not to disappoint my teacher.

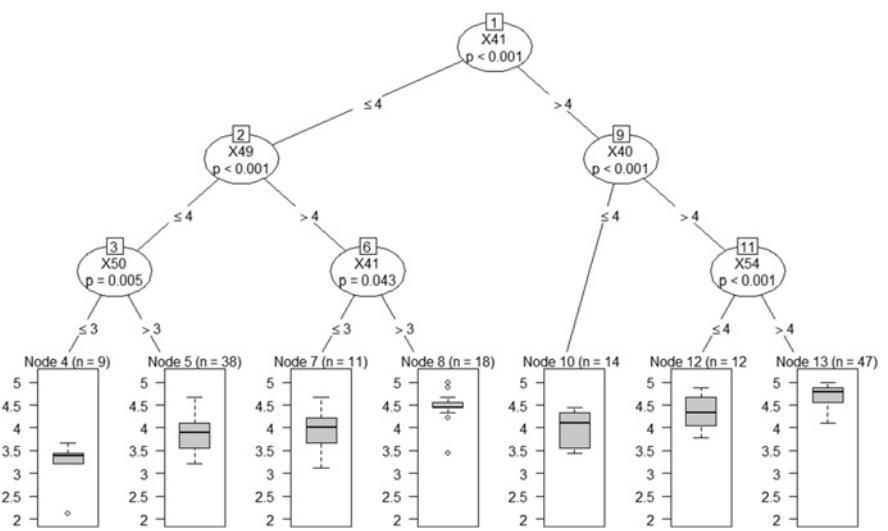

Figure 4: CIT for the item-based model of the data ($R^2 = 0.615$).

Perhaps the most interesting thing to be gleaned from this model is that the best predictors have changed: The top splits in the tree are induced by the subjects' experience of learning English, i. e. by whether they find the learning experience interesting (item 41) and whether they actually enjoy their engagement with the language (item 40). If both of these items are agreed to very affirmatively (i. e. in the subset comprising the three rightmost bins, or 73 participants), then item 54 further decides on the highest degrees of motivated behaviour found in the sample: The 47 students making up the rightmost bin all share the vision that their future job requires them to speak English, which in turn leads to extremely high degrees of motivation. At the other end of the spectrum, we find learners with scores below 4 on item 41 and a comparatively weaker vision of using the L2 effectively in the future (item 49). If this then co-occurs with a lack of vision to use the L2 in one's future profession (item 50), we get to the lowest scores of motivated learning behaviour in our data ($n = 9$, i. e. the leftmost bin). The fact that CIT models can highlight such interesting and complex interactions of factors is one of their great advantages over conventional regression models.

If we now grow a random forest with the item-specific predictors from Table 2, again by averaging over 3,000 permuted CITs, we obtain a similar but yet more complex result (Figure 5):

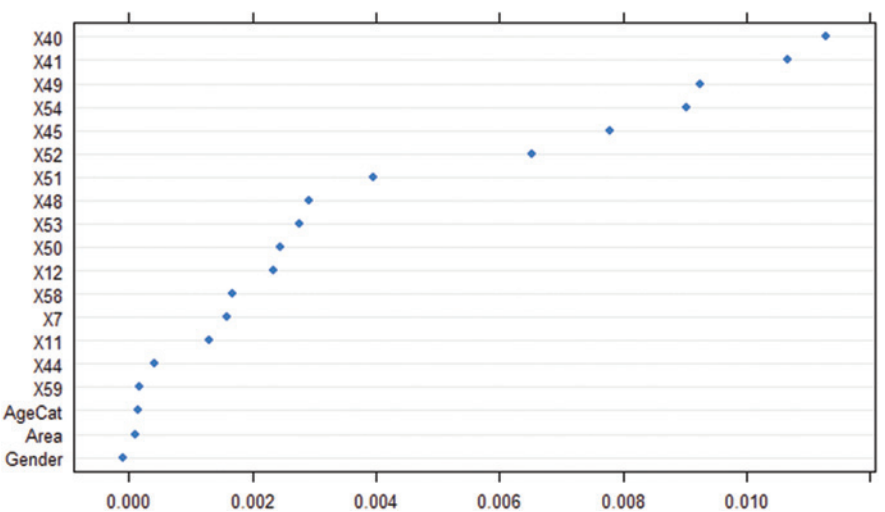

Figure 5: Conditional variable importance in the item-based model of the data ($R^2 = 0.669$).

Although nearly all of the predictors are technically significant, the strongest predictors are chiefly the ones that we also found in our CIT above, notably items 40 and 41, 49 and 54. These remain firmly in place across several random-forest computations, as do items 45 and 52: The latter two (which were not part of the CIT) both relate to using English as an international means of communication, either with people from other countries in general (item 45) or with international friends specifically (item 52). Therefore, these two items are conceptually very close and exert a similar influence on motivated learning behaviour, yet they had been taken to operationalize two different scales (cf. Table 2 again). This shows that the allocation of particular items to scales can be rather arbitrary, which is potentially problematic for scale-based analyses of the data. A similar problem arises with regard to items 53 and 54: In accordance with Kormos and Csizér (2008) and other previous studies (e. g. Ryan 2008), these items were here subsumed under the Ideal L2 Self scale, yet they are conceptually somewhat different from typical Ideal-Self items because they ask for whether English is seen by the learners as a *requirement* for the things they wish to do in the future (including their jobs). Therefore, while they do capture a certain vision of the participants as future L2 users, this vision could be argued to be related to the prevention of a negative outcome, just like failing a test or being considered an uneducated person.[20] To the extent that it is debatable whether these items relate to a promotion or a prevention focus, their allocation to the Ideal L2 Self scale or the Ought-to L2 Self scale is not completely straightforward. Now, recall that we dismissed the Ought-to L2 Self scale as unreliable in Section 3 above, but given that items 53 and 54 bear some flavours of this construct, and that, additionally, item 58 is also significant in Figure 5, it may be not be warranted to deny any influence from such Ought-to pressures on motivated learning behaviour. In our view, the item-based analysis presented here can bring such influences to the fore and thus usefully complements the traditional scale-based approach.[21] The same advantage can also be appreciated in relation to the strongest predictors in

20 Compare, for example, some of the 'prevention-focus' items in Taguchi et al. (2009), which are very similar in spirit to our items 53 and 54.

21 A different way of solving this problem is, of course, to work with a conceptually more stringent definition of the two constructs: In Kormos et al. (2011: 512), for example, the Ideal L2 Self is defined as "the learners' own internalized views of the value and importance of L2 learning", while the Ought-to Self captures "the external views of their environment". On this view, our items 53 and 54 are more clearly related to the Ideal L2 Self, but some of the items on Kormos and Csizér's (2008) Ought-to Self scale may, in fact, also have to be removed or reconceptualized.

Figures 4 and 5: Recall from above that items 40 and 41 did not form a reliable scale, neither in this combination ($\alpha = 0.64$) nor when we added items 39 and 42, as originally intended ($\alpha = 0.51$). If we had thus dismissed the scale as a whole, as the traditional procedure suggests, we would have missed some crucial predictors of motivated learning behaviour.[22]

In sum, the most stable and most influential predictors of our response variable in an item-based model of the data relate to the language-learning experience, which is in full accord with both Kormos and Csizér's (2008) regression models, as well as You and Dörnyei's (2016: 512) conclusion on Chinese learners of L2 English: "The desire to invest time and energy in language learning seems to be associated first and foremost with the evaluation of the learning process." Further effective motivating factors are constituted by items relating to the Ideal L2 Self (although these also include English as a *requirement* for specific future plans, not just a vision as a competent L2 speaker) and to international posture. This factor constellation is exactly in line with Kormos and Csizér's (2008) regression analysis, and our findings hence lend support to the cross-cultural validity of at least these aspects of the L2 Motivational Self System.

6 Concluding remarks

In this paper, we have provided the first study of the motivational disposition of Nepalese learners of L2 English. Expanding on previous questionnaire designs and analytical procedures, notably those in Kormos and Csizér (2008), we examined to what extent our data (i) provide evidence for the postulated latent dimensions underlying the questionnaire, (ii) mirror previous findings on gender-, region- and age-specific distributions, (iii) can be modelled by the theoretical components of the L2 Motivational Self System. On all three of these points, we find ourselves in broad agreement with other quantitative studies, including the ones on other Asian learner populations. Crucially, however, we argued for a more nuanced analysis of the data that emphasizes the role of specific items on the questionnaire rather than that of scales which abstract away from the actual data: Specifically, we noted that the allocation of items to scales can be problematic in the first place, as many of the

22 While it is true that our scale-based models in Figures 2 and 3 *did* contain the relevant scale, this scale comprised items 39 and 42, to the exclusion of 40 and 41, and according to Kormos and Csizér's (2008) procedure, two-item scales would have had to be discarded altogether. Either way, then, we would have missed potentially influential predictors.

scales tap into similar psychological domains and individual items may potentially be attributed to several scales. Secondly, if dimension-reduction techniques are used to eliminate internally heterogeneous scales from subsequent regression analyses, this runs the risk of overlooking that individual items from such scales may contribute significantly to explaining motivated learning behaviour. Finally, by using the means of the contributing items, a scale-based approach again abstracts from the data and may result in a different assessment of the relative importance of the respective scales in predicting motivated learning behaviour.

In order to sidestep these issues, we offered an item-based regression approach as a complement to the traditional scale-based procedure. In our specific case, we believe that the item-based model reflects the statistical structure of our data more faithfully. While it is true that both models lend strong support to the Ideal L2 Self as a major motivational antecedent, the item-based model can provide access to the internal structure of this construct and elucidate which of its subcomponents are the most effective catalysts of motivated learning behaviour. Furthermore, the scale-based model systematically underestimates the importance of the learning experience and instead highlights other factors (like instrumentality, or attitudes to the L2 community) that only play a minor role in the more 'direct' item-based approach. It is only with this latter procedure, therefore, that we can extract the strongest motivational antecedents, and that these square with the results of previous studies.

In conclusion, we have seen that our specific sample of Nepalese learners is exceptionally highly motivated to put efforts into learning English, regardless of the gender, age or regional affiliation of the participants. Our regression analyses showed that the previous learning experience is the major determinant in predicting motivated behaviour, followed by factors related to the learners' vision of themselves as competent L2 speakers, of future plans requiring an active command of English, and the openness to establishing international contacts by using English. The majority of these additional factors thus supports the Ideal L2 Self and the language-learning experience as central cornerstones of the L2 Motivational Self System. In keeping with previous studies, our assessment of the Ought-to L2 Self is more moderate, but our item-based model did suggest that a command of English may be seen as an externally-imposed requirement and hence as inducing motivated behaviour by a prevention focus.

At the very end of this paper, we should also point out some limitations of our study. Most importantly, it needs to be conceded that the present learner population is restricted in its size due to severe difficulties of accessing and recruiting educational institutions for the study. Therefore, the current sample cannot be taken to be representative of Nepal more widely. In particular, most of the questionnaires came from pupils and students from the urban Pokhara area, biasing the sample strongly

against the more traditional parts of the country. Moreover, since most of the schools in question charge tuition fees, the sample is further biased to students from comparatively affluent families (or students with scholarships). Finally, the unique characteristics of the remote school in our sample (cf. Section 2 again) render it impossible to generalize our findings to genuinely traditional areas of the country. Having said this, it is equally true that the uniqueness of this subgroup of our sample, with plain living conditions but access to formal education in English and L2 cultural products, only enabled us to conduct this kind of study: An investigation of rural areas lacking these characteristics would mandate a thorough revision of the questionnaire, and so we have actually been in a position to test the current survey instrument in a novel but suitable context.

Acknowledgements: Apart from the acknowledgements made directly in the text, we would like to express our gratitude to numerous people involved in the data-collection process as such. First of all, we thank all of the students and pupils who participated in this study and all of the teachers and principals that allowed us to introduce the project to them. Special thanks go to the staff of Himanchal Education Foundation, who accepted the second author of this paper as a volunteer and gave him the opportunity to conduct this study in Nangi Village. Dr Mahabir Pun and Chitra Bahadur Tilija Pun, HEF's primary representatives in Nepal, as well as Dr Debra Stoner and Dr Leonard Skov, volunteer advisers of HEF in Nebraska, constantly assisted in getting into contact with Nepali learners and teachers. Furthermore, this project would not have been possible without the help of Prakash Gautam, teacher at Enlightened International Academy and Sagarmatha Higher Secondary School, Pokhara, who translated the questionnaire and who allowed Gregor to visit his lessons and introduce the project to his students. We are also obliged to Alex Clark and Joceline Houdijk, two volunteers working in Pokhara, for taking some questionnaires to their classes and even feeding the data into the online version. Furthermore, we owe thanks to Prof. Dr Debendra Bahadur Lamichane, head of National Inventive Boarding School and lecturer at Thribuvan College, Pokhara, for encouraging over 50 students to participate in the project.

Appendix 1. Survey instrument

In this supplementary material, we provide the full questionnaire in a systematic form, where each scale is presented along with all of its constituent items. It goes without saying the participants of our study received a thoroughly randomized version of the questionnaire.

In the following lists, ...

- items marked with an *asterisk are those whose scores were inverted for the statistical analysis in order to maintain a consistent directionality in what the questions asked for (and in how they relate to motivated learning behaviour).
- items printed in *italics* were added to the present questionnaire (as compared to Kormos and Csizér 2008).

Integrativeness

1. How much would you like to become similar to the people who speak English?
2. How much do you like English?
3. How important do you think learning English is in order to learn more about the culture and art of its speakers?

Attitude towards the L2 community[23]

4. How much do you like the people who live in the United States?
5. How much do you like the people who live in the United Kingdom?
6. How much would you like to travel to the UK?
7. How much would you like to travel to the USA?
8. How much do you like meeting foreigners from English-speaking countries?

Instrumentality

9. How much do you think knowing English would help you if you travelled abroad in the future?
10. How much do you think knowing English would help you in the future?
11. How important do you think English is in the world these days?
12. Knowledge of English would make me a better educated person.

Cultural interest

13. How much do you like movies made in the United States?
14. How much do you like the pop music of the USA?
15. How much do you like the magazines made in the United States?

Vitality of the L2 community

16. How important is the United Kingdom in the world?
17. How rich and developed do you think the United Kingdom is?
18. How rich and developed do you think the United States is?
19. How important a role do you think the United States plays in the world?

23 In Ryan's (2008) questionnaire, this scale is called 'direct contact with L2 speakers', but the items are (nearly) identical.

Linguistic self-confidence

20. I am sure I will be able to learn a foreign language well.
21. Learning a foreign language is a difficult task.*
22. I think I am the type who would feel anxious if I had to speak to someone in a foreign language.*

Language-use anxiety

23. I feel uneasy speaking English with a native speaker.
24. I would get tense if a foreigner asked me for the way in English.
25. If there was an opportunity to meet an English speaker, I would feel nervous.
26. I am worried that native speakers of English would find my English strange.

Classroom anxiety

27. It embarrasses me to volunteer answers in our English class.
28. I get nervous when I am speaking in my English class.
29. I am afraid the other students will laugh at me when I speak English.
30. I always feel that the other students speak English better than I do.

Milieu

31. People around me tend to think that it is a good thing to know foreign languages.
32. Nobody really cares whether I learn English or not.*
33. For people where I live learning English is not really necessary.*
34. My parents consider foreign languages important school subjects.

Parental encouragement

35. My parents really encourage me to study English.
36. My parents encourage me to practise my English as much as possible.
37. My parents have stressed the importance English will have for me in my future.
38. My parents feel that I should really try to learn English.

Language-learning attitudes and experience

39. Learning English is really great.
40. I really enjoy learning English.
41. I find learning English really interesting.
42. I think that foreign languages are important school subjects.

International posture

43. Studying English will help me to understand people from all over the world (not just English-speaking countries).

44. If I could speak English well, I could get to know more people from other countries (not just English-speaking countries).
45. I would like to be able to use English to communicate with people from other countries.
46. In the future, I imagine myself working with people from other countries (not just English speaking countries).
47. In the future, I really would like to communicate with foreigners.

Ideal L2 Self

48. I like to think of myself as someone who will be able to speak English.
49. If my dreams come true, I will use English effectively in the future.
50. Whenever I think of my future career, I imagine myself being able to use English.
51. When I think about my future, it is important that I use English.
52. I can imagine speaking English with international friends.
53. The things I want to do in the future require me to speak English.
54. The job I imagine having in the future requires that I speak English well.

Ought-to L2 Self

55. If I fail to learn English I'll be letting other people down.
56. Learning foreign languages makes me fear that I will feel less Nepali because of it.*
57. Learning English is necessary because it is an international language.
58. For me to be an educated person I should be able to speak English.
59. *I study hard in order not to disappoint my teacher.*

Self efficacy

60. *I am responsible for the quality of my English.*
61. *I am not responsible for the quality of my English.**
62. *The quality of my English does not depend on factors that I cannot control (such as the quality of my teachers, lessons and learning material).*
63. *The quality of my English depends on factors that I cannot control (such as the quality of my teachers, lessons and learning material).**
64. *The quality of my English depends how much effort I put into my studies.*
65. *No matter how much I study, it does not improve my English.**
66. *If I want to speak English better, I have to work harder in the future.*
67. *If I continue studying English as I do now, my English will be very good.*

Perceived L2 difficulty

68. *I think English is more difficult than Nepali.*
69. *I like English but it is so hard to learn.*

70. *I think English is a very difficult language.*
71. *I don't like English because it is so difficult.*

Motivated learning behaviour

72. I am willing to work hard at learning English.
73. I am determined to push myself to learn English.
74. I can honestly say that I am really doing my best to learn English.
75. It is very important for me to learn English.
76. Learning English is one of the most important aspects in my life.
77. If an English course was offered in the future, I would like to take it.
78. If I could have access to English-speaking radio, I would try to listen to it often.
79. When I hear an English song on the radio, I listen carefully and try to understand all the words.
80. I always look forward to our English classes.
81. *Sometimes I study more than is expected of me.*
82. *I like to improve my English outside of the classroom.*

Appendix 2. Statistical analysis

In this supplementary material, we detail the analytical procedures of our study. All analyses were performed in the open-software *R*, version 3.2.3 (R Development Core Team 2015).

Dimension-reduction procedures

As explained in the main text, we used principal component analysis (PCA) in conjunction with factor analysis (FA) for testing the dimensionality of the purported scales. For each scale in question, we began the analysis by checking three crucial prerequisites for both PCA and FA: First, we applied the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy in order to test whether, given a particular set of items to be factorized, our sample of observations ($n_{max} = 149$) was sufficient to apply factor analysis in the first place.[24] Following recommendations by Kaiser (1974), we took KMO values > 0.5 to be

24 Since no *R* package currently implements KMO calculations, we are grateful to G. Jay Kerns for making a KMO function freely available, cf. http://tolstoy.newcastle.edu.au/R/help/05/12/17233.html (accessed 21 February 2016).

acceptable; lower values indicated that no PCA/FA could be performed (without, for example, removing any of the items in question, cf. also Backhaus et al. 2016: 398). Secondly, the correlational structure of the respective items was tested, in two complementary ways: Bartlett's test of sphericity (Bartlett 1951), implemented in the *R* function `cortest.bartlett()` in the 'psych' package, determines whether there is a minimal correlational structure in the data (by comparing the data to a so-called identity matrix without any correlations). The opposite scenario, i. e. that the items correlate too highly, is as problematic for FA as such extreme multicollinearity is for multiple regression models, because it becomes impossible to gauge the unique contributions of highly correlated items to a given factor. This scenario was tested by calculating the so-called determinant, using the `det()` function from the *R* base distribution; the determinant value should be substantially greater than 0.00001 (since perfect correlation, so-called singularity, is indicated by a determinant of zero, cf. Field et al. 2012: 771).

For the PCAs, we used the `PCA()` function from the 'FactoMineR' package. Missing values in our data were imputed by the function `imputePCA()` from the 'missMDA' package (cf. Josse and Husson 2012 for technical details). For determining the optimal number of principal components, we inspected the eigenvalues of each component, using scree plots (cf. Catell 1966) in conjunction with Kaiser's (1960) criterion of choosing components with eigenvalues greater than 1. The ideal scenario in our case is, of course, one in which a single principal component emerges and where, following Stevens' (2002) guidelines for our sample size of $100 < n < 200$, each item has a loading of > 0.4 onto the component in question.

For the factor analysis, we used the function `factanal()` from the *R* base distribution. In each case, we took the smallest number of components suggested as adequate by the PCA (i. e. ideally just one) and tested whether this yields a satisfactory factor solution in FA. In order to improve interpretability with multi-component solutions, we employed 'promax' as an oblique method for factor rotation.[25] To inspect the factor loadings, we examined the pattern matrix and again took Stevens' > 0.4 criterion as the benchmark for the contribution of an item to a factor. Finally, we also used the maximum-likelihood residual analysis implemented in `factanal()`, which indicates whether the

25 As explained by Field et al. (2012: 765–767), oblique rotation should be preferred to orthogonal rotation methods like 'varimax' if one has theoretical reasons to believe that the factors in question might be correlated with one another. This is often the case in psychometric surveys (cf. also Matsunaga 2010: 107 on this point), and in the present study in particular (where many scales are closely related, and even more so the factors that emerge from the items taken to belong to the same scale).

number of factors specified is sufficient in order to account for the variation in the participating items.[26]

The last step of the analysis was to assess the reliability of the resulting scales, applying Cronbach's alpha (Cronbach 1951). Following standard recommendations in the literature (e. g. also Dörnyei 2010b: 95 on questionnaire data), we used $\alpha \geq 0.7$ as the critical threshold for the reliability of a scale. In Appendix 1, where we provide the full questionnaire, we also indicate which items were reversed for the reliability analysis.

In the following overview, we turn to each scale from the paper individually again and provide the central results from the PCA, FA and reliability analysis. Unless specified otherwise, it can be taken for granted that both the Bartlett criterion and the determinant yielded acceptable results.

- **Integrativeness** (items 1–3): KMO = 0.57. PCA suggests that a single-component solution may be acceptable, but this is not corroborated by FA (item 3 showed a loading < 0.4 and a single-factor solution is not judged sufficient). $\alpha_{(1-3)} = 0.48$, $\alpha_{(1-2)} = 0.42$. We thus obtain a heterogeneous and unreliable scale, even if shorted to the more homogeneous items 1 and 2.
- **Attitudes towards the L2 speakers and the community** (items 4–8): KMO = 0.7. PCA suggests that a single-component solution is perfectly adequate. In FA, a single-factor solution is not judged sufficient by the significance test, but all items show loadings > 0.4 on the factor. The variance accounted for (and the solution as a whole) can be improved if item 8 is removed, and the final scale reported in the paper is reliable at $\alpha_{(4-7)} = 0.71$.
- **Instrumentality** (items 9–12): An initial analysis showed that item 12, which (in contrast to Kormos and Csizér 2008) we had reallocated to this scale because of its instrumental flavour, introduced severe heterogeneity, as it did not correlate with the other items on the scale. Consequently, we abandoned it from the scale and operated with items 9–11 only. KMO = 0.64. PCA then suggested a satisfactory one-component solution. FA's significance test did not consider a single factor to be sufficient, but again all items had acceptable loadings of > 0.4 on the factor. The scale is borderline reliable ($\alpha_{(9-11)} = 0.68$) and we hence retained it in further analyses.
- **Cultural interest** (items 13–15): KMO = 0.57. PCA suggests that a single-component solution may be acceptable, but this is not corroborated by FA (item 3 showed a loading < 0.4 and a single-factor solution is not judged

26 In practice, this test can yield the result that the number of factors is not sufficient, despite the fact that the factor solution is actually good in terms of the loadings and other parameters. For this reason, we sometimes neglect the results from this test if all other indicators point to a satisfactory one-dimensional factor structure.

sufficient). $\alpha_{(13-15)} = 0.57$. The scale is thus both heterogeneous and unreliable.

- **Vitality of the L2 community** (items 16–19): KMO = 0.72. PCA suggests a clearly one-dimensional structure with high loadings of all items. FA with one factor confirms loadings > 0.4, but fails the significance test. In view of the strong signals from all other criteria, however, the scale was included in its original composition. This is confirmed by reliability analysis ($\alpha_{(16-19)} = 0.78$).
- **Linguistic self-confidence** (items 20–22): In this case, the Bartlett test yields a non-significant result ($p = 0.08$), and KMO = 0.44. Therefore, the correlational structure of these items is so weak that not even the basic prerequisites for PCA/FA are fulfilled.
- **Language use anxiety** (items 23–26): KMO = 0.73. Both PCA and FA suggest a one-dimensional structure with high loadings of all items onto the single factor. This monofactorial solution also passes FA's significance test and is reliable at $\alpha_{(23-26)} = 0.74$.
- **Classroom anxiety** (items 27–30): KMO = 0.65. An initial analysis showed that item 30 makes the scale inconsistent and leads to an unacceptable alpha value. Removing this item returns a single-component structure in PCA and a monofactorial solution in FA with all loadings > 0.4. This scale is reliable at $\alpha_{(27-30)} = 0.73$.
- **Milieu** (items 31–34): This scale violates some crucial prerequisites for PCA/FA, i. e. Bartlett's test is non-significant ($p = 0.08$) and KMO = 0.49. Even if we remove item 32 because of its large amount of NAs, the results do not improve significantly. Not surprisingly, then, the scale is also completely unreliable ($\alpha_{(31-34)} = 0.09$).
- **Parental encouragement** (items 35–38): KMO = 0.77. Although both PCA and FA suggest a satisfactory one-dimensional solution, some improvement in the factor loadings and in the variance explained by a single factor can be achieved if item 36 is removed. The resulting scale is reliable at $\alpha_{(35,37,38)} = 0.74$.
- **Language-learning attitudes** (items 39–42): KMO = 0.51. Both PCA and FA suggest a two-component solution with complex cross-loadings and an overall unreliable composition ($\alpha_{(39-42)} = 0.51$). As explained in the paper, a subscale of items 39 and 42 achieves a much higher degree of reliability ($\alpha_{(39,42)} = 0.7$) and was hence retained in further analyses.
- **International posture** (items 43–47): Item 45 had to be removed due to a large number of NAs. For the remaining items, KMO = 0.53. PCA suggests that a single-component solution might be possible (though certainly not optimal), and all signals of a FA clearly speak against this. Removing individual

items does not improve this situation, and the scale is not reliable ($\alpha_{(43-47)} = 0.55$).

- **Ideal L2 Self** (items 48–54): KMO = 0.76. Both PCA and FA suggest a two-dimensional construct, but since an outstanding first component emerges in PCA, with high loadings of all items, and since a monofactorial solution is still reliable ($\alpha_{(48-54)} = 0.71$), we had sufficient reason to believe that this construct is valid enough (and certainly more homogeneous than truly diverse scales). Note that removing items with the lowest loadings does not improve the residual diagnostics in FA, nor does it lead to a more reliable alpha value.
- **Ought-to L2 Self** (items 55–59): KMO = 0.59. Both PCA and FA suggest a complex (at least two-factor) structure with many cross-loadings, and an insufficient $\alpha_{(55-59)}$ of −0.52.
- **Self-efficacy** (items 60–67): KMO = 0.54. Both PCA and FA suggest a very heterogeneous composite of three factors. As far as reliability is concerned, neither the scale in its original composition ($\alpha_{(60-67)} = 0.06$) nor in a version reduced to the most strongly correlating items ($\alpha_{(60,\ 64,\ 67)} = 0.52$) reaches the criterion level.
- **Perceived L2 difficulty** (items 68–71): KMO = 0.69. Both PCA and FA suggest a satisfactory one-component solution, as all items show high loadings on the first dimension and even the residual analysis shows that a monofactorial analysis is sufficient. The scale is reliable at $\alpha_{(68-71)} = 0.7$.
- **Motivated learning behaviour** (items 72–82): A first round of analysis showed that the two items we added to the original scale (i. e. 81 and 82) introduced heterogeneity. Without them, KMO = 0.83. PCA suggests that a one-dimensional solution may be warranted, with high loadings of all items on the first dimension (similarly to the Ideal L2 Self above). A similarly good loading structure emerges in a monofactorial FA (even though this does not pass the residual test). The scale is reliable at $\alpha_{(72-80)} = 0.83$.

Statistical parameters of reliable scales

This part of the appendix relates to Section 4 of the paper. The relevant data can be found in Table 3 on the next page, which has to be read as follows: For each of the reliable scales listed in the first column, we provide comparative data on three different subsamples, i. e. female/male, remote/urban and pupil/student. We first list the mean and standard deviation for each part of the respective subsample (e. g. female versus male) and then compare the difference in their average scores by means of non-parametric Wilcoxon rank-sum tests (excluding missing values on a pairwise basis).

Table 3: Statistical properties of the reliable scales.

Scale	Items	Sample contrasts	Statistics of sample contrasts						Correlation with motivated learning behaviour							
			Mean		SD		*W*	*p(W)*	*ρ*	*p(ρ)*	*ρ*		*p(ρ)*		*z*	*p(z)*
Attitudes to the L2 community	4–7	female:male	3.91	3.90	0.6	0.77	2474	0.636			0.2	0.64	0.08	<0.001	3.20	0.001**
		remote:urban	3.70	4.01	0.91	0.57	1912	0.096	0.46	<0.001***	0.486	0.468	<0.001	<0.001	0.13	0.897
		pupils:students	3.88	4.00	0.72	0.64	2238.5	0.596			0.463	0.512	<0.001	<0.001	−0.35	0.726
Instrumentality	9–11	female:male	4.67	4.50	0.50	0.59	2951	0.119			0.376	0.447	0.001	<0.001	0.50	0.617
		remote:urban	4.77	4.49	0.52	0.57	3113	<0.001***	0.42	<0.001***	0.549	0.382	<0.001	<0.001	1.16	0.246
		pupils:students	4.48	4.77	0.60	0.43	1760.5	0.007**			0.425	0.264	<0.001	0.038	1.01	0.313
Vitality of the L2 community	16–19	female:male	4.07	4.11	0.65	0.70	2448.5	0.565			0.017	0.267	0.445	0.011	1.51	0.131
		remote:urban	3.90	4.16	0.72	0.64	1767.5	0.023*	0.157	0.028*	0.205	0.210	0.091	0.016	−0.03	0.976
		pupils:students	4.08	4.08	0.68	0.65	2378	0.972			0.131	0.280	0.093	0.030	−0.85	0.395
Language use anxiety	23–26	female:male	2.98	2.91	1.01	0.98	2737	0.563			−0.366	−0.20	<0.001	0.044	0.15	0.294
		remote:urban	2.90	2.97	1.12	0.92	2213.5	0.689	−0.223	0.003**	−0.602	−0.028	<0.001	0.387	−3.61	<0.001***
		pupils:students	3.00	2.84	1.09	0.68	2655	0.239			−0.286	0.151	0.002	0.843	−2.45	0.014*
Classroom anxiety	27–29	female:male	2.75	2.57	1.08	0.99	2810.5	0.381			−0.389	−0.251	0.016	<0.001	0,91	0.363
		remote:urban	2.72	2.64	1.15	0.98	2392	0.733	−0.282	<0.001***	−0.633	−0.128	<0.001	0.096	−3.34	<0.001***
		pupils:students	2.75	2.47	1.05	0.96	2739	0.127			−0.382	0.068	<0.001	0.673	−2.58	0.01*
		remote:urban	4.23	3.94	0.79	0.81	2859.5	0.021*	0.487	<0.001***	0.419	0.431	0.002	0.003	−0.08	0.936
		pupils:students	4.03	4.02	0.80	0.85	2352.5	0.9469			0.543	0.328	<0.001	0.013	1,47	0.142
Learning attitudes/ experience	39,42	female:male	4.18	4.12	0.74	0.98	2532.5	0.809			0.213	0.515	0.004	<0.001	2.08	0.038*
		remote:urban	3.63	4.34	0.89	0.79	1189.5	<0.001***	0.357	0.005**	0.443	0.429	0.001	<0.001	0.10	0.920
		pupils:students	4.14	4.09	0.82	1.00	2351	0.941			0.311	0.559	<0.001	<0.001	−1,7	0.089
Ideal-L2-Self	48–54	female:male	4.38	4.24	0.43	0.50	3022.5	0.084			0.743	0.751	<0.001	<0.001	0.11	0.912
		remote:urban	4.36	4.29	0.54	0.44	2577	0.265	0.751	<0.001***	0.863	0.677	<0.001	<0.001	2.61	0.009**
		pupils:students	4.23	4.49	0.47	0.42	1559	<0.001***			0.786	0.684	<0.001	<0.001	1.22	0.225
Perceived L2 difficulty	68–71	female:male	2.70	2.67	0.82	0.93	2651.5	0.813			−0.458	−0.171	<0.001	0.075	1.89	0.059*
		remote:urban	2.79	2.65	0.85	0.91	2514	0.396	−0.214	0.005**	−0.565	−0.155	<0.001	0.057	−2.62	0.009**
		pupils:students	2.84	2.36	0.86	0.88	3131	0.002**			−0.365	0.158	<0.001	0.853	−2.97	0.003**
Motivated learning behaviour	72–80	female:male	4.29	4.14	0.47	0.59	2932	0.174								
		remote:urban	4.27	4.20	0.60	0.51	2582.5	0.257								
		pupils:students	4.14	4.41	0.54	0.48	1647	0.003**								

In the right-hand part of the table, we then display the (non-parametric) correlation coefficient Spearman's ρ (*rho*) for the correlation of the scale in question with the scale of motivated learning behaviour, i. e. the mean value of items 72–80. In testing these correlations for significance, we generally employed one-tailed tests because our hypotheses were clearly directional (e. g. we supposed that an Ideal L2 Self would correlate positively with motivated learning behaviour). However, for testing such correlations in the subsamples (e. g. differences between male and female participants), two-tailed tests were used.

The last two columns of the table feature a test statistic z and its associated p-value. z was obtained by applying Fisher's r-to-z transformation, which allows comparing whether two correlation coefficients (e. g. that of the rural and the urban subsample) differ significantly from one another.[27]

Regression analyses

As is explained in the main text, our specific data set (with many predictors relative to the number of observations, heavy deviations from normality, potentially correlated predictors, etc.) suggests employing a specific non-parametric regression procedure known as conditional inference trees (CIT) and random forests (RF). The gist of these techniques is described in the main text, and we elaborate on them here and add some specifications of the software and algorithms we drew on.

Individual CITs were produced by using the `ctree()` function in the 'party' package (cf. Strobl et al. 2009a), which has an inbuilt 'stop criterion' when no significant association between a predictor and the response variable can be detected anymore, thus preventing further splits of the tree. The function works on the basis of an unbiased algorithm, which (in contrast to other implementations like `rpart()`) does not favour variables with many distinct levels (such as continuous or multilevel categorical predictors over binary ones). The function is thus particularly suitable to our specific mix of predictors.

RFs are so-called ensemble methods in which many individual trees are grown from random subsets of the data. For each of the trees involved, the whole sample is randomly divided into a learning set (the so-called 'in-bag' observations) and a test set (the so-called 'out-of-bag' observations), and only a random subset of the predictor variables is used to generate the tree. In this way,

27 We are grateful to Richard Lowry for providing a freely available web interface for performing the relevant calculations, cf. http://vassarstats.net/rdiff.html (accessed 20 March 2016).

a truly diverse set of trees is grown, and by averaging over these trees, the relative importance of each predictor can be assessed more reliably than in a single tree model for the data (since a whole ensemble of trees is less sensitive to idiosyncratic properties of the specific sample at hand). The relative weight of a predictor is gauged by a rather complex procedure of permutation (cf. Breiman 2001). The basic idea is that the values of the predictor variable in question are shuffled so that there is no association with the response variable anymore; if this permuted version of the predictor weakens the performance of the model as compared to using the original, unpermuted predictor, one can conclude that the predictor in question is essential for modelling the response variable (cf. also Tagliamonte and Baayen 2012: 160–161 for a digestible summary of the technicalities).

For growing random forests, we used the `cforest()` function of the 'party' package.[28] For the reasons mentioned above, we used the default option `controls = cforest_unbiased` to ensure unbiased treatment of different types of variables in each tree. Furthermore, in assessing the relative impact of the predictors, we extracted the *conditional* permutation variable importance, using the `varimp(obj, conditional = T)` routine. This procedure, proposed in Strobl et al. (2008), specifically ensures that the importance of correlated predictors is not overestimated and thus reduces the potential problem of collinearity. In growing the random forest, we used the default `mtry = 5` setting (i. e. with five randomly preselected predictor variables at each split of each tree) but with a considerably larger number of trees than specified in the default settings (3,000 instead of 500), in order to do justice to the rather high number of predictor variables and to make the results more reliable (cf. Strobl et al. 2009b: 343). Following recommended practices, we tried several different parameter settings (including varying random seeds), but obtained stable results as far as the relative variable importance of the strongest predictors was concerned (while the relative ranking of less influential predictors changed across the various runs).

For determining the statistical significance of a predictor in the random forest solution, we followed Strobl, Malley and Tutz (2009: 343), who suggest that "all variables with importance that is negative, zero, or positive but with a value that lies in the same range as the negative values can be excluded from further exploration. The rationale for this rule of thumb is that the importance of

28 Note that, for the item-based analysis in Section 5.2, missing values on individual items had to be imputed (i. e. replaced by the median value for the item in question) before an RF analysis could be run. To this end, we used the function `na.roughfix()` from the 'randomForest' package (cf. also Liaw and Wiener 2002).

irrelevant variables varies randomly around zero. Therefore, positive variation of an amplitude comparable to that of negative variation does not indicate an informative predictor variable, whereas positive values that exceed this range may indicate that a predictor variable is informative." More specifically, we took the absolute value of the most negative predictor as a threshold for the significance of the positive predictors.

The goodness-of-fit of CITs and RFs can be obtained in a similar way as for multiple linear regression models, i. e. by using the regular formula for the coefficient of determination: $R^2 = 1 - SS_{res}/SS_{tot}$, where SS_{res} denotes the residual sum of squares and SS_{tot} the total sum of squares (cf. Field et al. 2012: 250). This procedure capitalizes on the fact that, in both CITs and RFs, one can compare the predicted and the observed values of the response variable (based on a single model in CITs and the average of many models in RFs). What we observe in our data with regard to R^2 (compare Figures 2 and 3, or 4 and 5) is rather typical: The goodness-of-fit of a single CIT is usually slightly worse than the predictive accuracy of a whole ensemble of trees, i. e. of a RF, especially if the latter is grown from a very large number of trees.

References

Apple, Matthew T., Dexter Da Silva & Terry Fellner (eds.). 2016. *L2 selves and motivations in Asian contexts*. Clevedon: Multilingual Matters.

Backhaus, Klaus, Bernd Erichson, Wulff Plinke & Rolf Weiber. 2016. *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*, 14th edn. Berlin & Heidelberg: Springer.

Bartlett, Maurice S. 1951. The effect of standardization on a chi square approximation in factor analysis. *Biometrika* 38. 337–344. Doi:10.2307/2332580.

Bong, Mimi & Richard E. Clark. 1999. Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist* 34(3). 139–153. Doi:10.1207/s15326985ep3403_1.

Boo, Zann, Zoltán Dörnyei & Stephen Ryan. 2015. L2 motivation research 2005–2014: Understanding a publication surge and a changing landscape. *System* 55. 145–157. Doi:10.1016/j.system.2015.10.006.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45(1). 5–32. Doi:10.1023/A:1010933404324.

Carlson, Rae. 1965. Stability and change in the adolescent's self-image. *Child Development* 36. 659–666. Doi:10.2307/1126912.

Catell, Raymond B. 1966. The scree test for the number of factors. *Multivariate Behavioral Research* 1. 245–276. Doi:10.1207/s15327906mbr0102_10.

Cronbach, Lee J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16. 297–334. Doi:10.1007/bf02310555.

Csizér, Kata & Zoltán Dörnyei. 2005. The internal structure of language learning motivation and its relationship with language choice and learning effort. *The Modern Language Journal* 89(1). 19–34. Doi:10.1111/j.0026-7902.2005.00263.x.

Csizér, Kata & Judit Kormos. 2009. Learning experiences, selves and motivated learning behavior: A comparative analysis of structural models for Hungarian secondary and university learners of English. In Zoltán Dörnyei & Ema Ushioda (eds.), *Motivation, language identity and the L2 Self*, 98–119. Clevedon: Multilingual Matters.

Dörnyei, Zoltán. 2005. *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Dörnyei, Zoltán. 2009. The L2 motivational self system. In Zoltán Dörnyei & Ema Ushioda (eds.), *Motivation, language identity and the L2 Self*, 9–42. Clevedon: Multilingual Matters.

Dörnyei, Zoltán. 2010b. *Questionnaires in second language research: Construction, administration, and processing*, 2nd edn. With contributions from Tatsuya Taguchi. New York & London: Routledge.

Dörnyei, Zoltán. 2010a. Researching motivation: From integrativeness to the Ideal L2 Self. In Susan Hunston & David Oakey (eds.), *Introducing applied linguistics: Concepts and skills*, 74–83. London: Routledge.

Dörnyei, Zoltán, Christine Muir & Zana Ibrahim (2014). Directed Motivational Currents: Energising language learning by creating intense motivational pathways. In David Lasagabaster, Aintzane Doiz & Juan Manuel Sierra (eds.), *Motivation and foreign language learning: From theory to practice*, 9–29. Amsterdam & Philadelphia: John Benjamins.

Field, Andy, Jeremy Miles & Zoe Field. 2012. *Discovering statistics using R*. Los Angeles etc.: Sage.

Gardner, Robert C. 1985. *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.

Gardner, Robert C. 2001. Integrative motivation and second language acquisition. In Zoltán Dörnyei & Richard Schmidt (eds.), *Motivation and second language acquisition*, 1–19. Honolulu, HI: University of Hawai'i Press.

Gass, Susan M. & Larry Selinker. 2008. *Second language acquisition: An introductory course*, 3rd edn. Mahwah, NJ: Lawrence Erlbaum.

Gobel, Peter, Siew Ming Thang & Setsuko Mori. 2016. Revisiting studies on causal attribution in ESL/EFL contexts: Toward an alternative model. In Matthew T. Apple, Dexter Da Silva & Terry Fellner (eds.), *L2 selves and motivations in Asian contexts*, 11–28. Clevedon: Multilingual Matters.

Henry, Alastair. 2011. Gender differences in L2 motivation: A reassessment. In Samuel A. Davies (ed.), *Gender gap: Causes, experiences and effects*, 81–101. New York: Nova Science.

Henry, Alastair & Christina Cliffordson. 2013. Motivation, gender, and possible selves. *Language Learning* 63(2). 271–295.

Higgins, E. Tory. 1987. Self-discrepancy: A theory relating self and affect. *Psychological Review* 94. 319–340. Doi:10.1037//0033-295x.94.3.319.

Higgins, E. Tory. 1998. Promotion and prevention: Regulatory focus as a motivational principle. *Advances in Experimental Social Psychology* 30. 1–46. Doi:10.1016/s0065-2601(08)60381-0.

Jonkisz, Ewa, Helfried Moosbrugger & Holger Brandt. 2011. Planung und Entwicklung von Tests und Fragebogen. In Helfried Moosbrugger & Augustin Kelava (eds.), *Testtheorie und Fragebogenkonstruktion*, 2nd edn., 27–74. Berlin & Heidelberg: Springer.

Josse, Julie & François Husson. 2012. Handling missing values in exploratory multivariate data analysis methods. *Journal de la Société Française de Statistique* 153(2). 79–99.

Kaiser, Henry F. 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 20. 141–151. Doi:10.1177/001316446002000116.
Kaiser, Henry F. 1974. A second-generation little jiffy. *Psychometrika* 35. 401–415. Doi:10.1007/bf02291817.
Kormos, Judit & Kata Csizér. 2008. Age-related differences in motivation of learning English as a foreign language: Attitudes, selves, and motivated behavior. *Language Learning* 58. 327–355. Doi:10.1111/j.1467-9922.2008.00443.x.
Kormos, Judit, Kata Csizér, Adrienn Menyhárt & Dóra Török. 2008. 'Great Expectations': The motivational profile of Hungarian English language students. *Arts and Humanities in Higher Education* 7(1). 65–82. Doi:10.1177/1474022207084884.
Kormos, Judit & Thom Kiddle. 2013. The role of socio-economic factors in motivation to learn English as a foreign language: The case of Chile. *System* 41. 399–412. Doi:10.1016/j.system.2013.03.006.
Kormos, Judit, Thom Kiddle & Kata Csizér. 2011. Systems of goals, attitudes, and self-related beliefs in second-language-learning motivation. *Applied Linguistics* 32(5). 495–516.
Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam & Philadelphia: John Benjamins.
Liaw, Andy & Matthew Wiener. 2002. Classification and regression by randomForest. *R News* 2(3). 18–22.
Lohmann, Arne. 2013. Is tree hugging the way to go? Classification trees and random forests in linguistic study. *VIEWS: Vienna Working Papers in Linguistics* 22. 1–16. http://anglistik.univie.ac.at/views (accessed 3 March 2016).).
Markus, Hazel R. & Paula Nurius. 1986. Possible selves. *American Psychologist* 41. 954–969. Doi:10.1037//0003-066x.41.9.954.
Matsunaga, Masaki. 2010. How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research* 3(1). 97–110.
Noels, Kimberley A. 2003. Learning Spanish as a second language: Learners' orientations and perceptions of their teachers' communication style. *Language Learning* 53. 97–136. Doi:10.1111/1467-9922.53225.
R Development Core Team. 2015. *R: A language and environment for statistical computing. Version 3.2.3*. Vienna, Austria: R Foundation for Statistical Computing.
Ruvolo, Ann P. & Hazel R. Markus. 1992. Possible selves and performance: The power of self relevant imagery. *Social Cognition* 10(1). 95–124. Doi:10.1521/soco.1992.10.1.95.
Ryan, Stephen (2008). *The Ideal L2 Selves of Japanese learners of English*. University of Nottingham dissertation.
Stevens, James P. 2002. *Applied multivariate statistics for the social sciences*, 4th edn. Hillsdale, NJ: Lawrence Erlbaum.
Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9. 307. Doi:10.1186/1471-2105-8-25.
Strobl, Carolin, Torsten Hothorn & Achim Zeileis. 2009a. Party on!. *The R Journal* 1(2). 14–17.
Strobl, Carolin, James Malley & Gerhard Tutz. 2009b. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4). 323–348. Doi:10.1037/a0016973.
Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 34. 135–178. Doi:10.1017/s0954394512000129.

Taguchi, Tatsuya, Michael Magid & Mostafa Papi. 2009. The L2 motivational self system amongst Chinese, Japanese and Iranian learners of English: A comparative study. In Zoltán Dörnyei & Ema Ushioda (eds.), *Motivation, language identity and the L2 Self*, 66–97. Clevedon: Multilingual Matters.

Ushioda, Ema. 2001. Language learning at university: Exploring the role of motivational thinking. In Zoltán Dörnyei & Richard Schmidt (eds.), *Motivation and second language acquisition*, 91–124. Honolulu, HI: University of Hawai'i Press.

Wiechmann, Daniel. 2011. Exploring probabilistic differences between genetically related languages. *Languages in Contrast* 11(2). 193–215. Doi:10.1075/lic.11.2.03wie.

Yashima, Tomoko, Lori Zenuk-Nishide & Kazuaki Shimizu. 2004. The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning* 54(1). 119–152. Doi:10.1111/j.1467-9922.2004.00250.x.

You, Chenjing & Zoltán Dörnyei. 2016. Language learning motivation in China: Results of a large-scale stratified survey. *Applied Linguistics* 37(4). 495–519. Doi:10.1093/applin/amu046.

Zimmerman, Barry J. 1995. Self-efficacy and educational development. In Albert Bandura (ed.), *Self-efficacy in changing societies*, 202–231. Cambridge: Cambridge University Press.

Zimmerman, Barry J., Albert Bandura & Manuel Martinez-Pons. 1992. Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal* 29(3). 663–676. Doi:10.2307/1163261.