View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Crossref

Haggerty and Fox *Language Testing in Asia* (2015) 5:11
DOI 10.1186/s40468-015-0018-0

Language Testing in Asia
a SpringerOpen Journal

RESEARCH Open Access

Raising the bar: language testing experience and second language motivation among South Korean young adolescents

John F Haggerty[1*] and Janna Fox[2]

* Correspondence: john.haggerty@alumni.ubc.ca
[1]Department of Language and Literacy Education, University of British Columbia, Vancouver, BC, Canada
Full list of author information is available at the end of the article

Abstract

Drawing on second language (L2) motivation constructs modelled on Dörnyei's (2009) L2 Motivational Self System, this study explores the relationship between language testing experience and the motivation to learn English among young adolescents (aged 12–15) in South Korea.
A 40-item questionnaire was administered to middle-school students ($N = 341$) enrolled in a private language school (*hakwan*). Exploratory factor analysis (EFA) identified five salient L2 motivation factors. These factors were compared to four learner-background characteristics: gender, grade level, L2 test-preparation time, and experience taking a high-stakes university-level language test.
The results suggest that second language motivation, based on the L2 motivation factors identified as most salient in this educational context, was significantly associated with the amount of time spent preparing for language tests and experience taking a high-stakes language test intended primarily for university-entrance purposes.
Young South Korean adolescent learners' testing experiences and their motivation to learn English are discussed in relation to the social consequences of test use and ethical assessment practices.

Keywords: L2 motivation; Language testing; High-stakes testing; Ethical assessment; L2 Motivational Self System; Exploratory factor analysis

Background

The laudable goals of improving educational standards, promoting educational accountability, and encouraging learners to achieve have typically been accompanied by the use of standardized tests to measure progress and/or achievement. In fact, such tests are often viewed as the primary means to achieve such ends (Elwood, 2013; Qi 2007). Large-scale tests used for such purposes are often intended to be (or tend to become) high-stakes for test takers and/or other stakeholders (Stobart and Eggen, 2012). As Qi (2007) has pointed out, high-stakes tests "possess power to exert an expected influence on teaching and learning because of the consequences they bring about" (p. 52). There is now a general consensus that high-stakes testing practices have considerable influence on many teaching and learning environments around the

Springer

© 2015 Haggerty and Fox. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

world (Ross, 2008; Stobart and Eggen, 2012); however, highly complex dynamics make it difficult to establish causal relationships (Cheng, 2008, p. 358).

Concern has also been expressed about the impact of high-stakes language testing on both young (Kindergarten-grade 6) and adolescent (grades 7–12) language learners. For example, in Canada, government-mandated standards-driven tests have created concern over their potential to worsen high-school dropout rates for second language (L2) learners (Fairbairn and Fox 2009; Fox and Cheng 2007; Murphy Odo, 2012). Similarly in the U.S., the far-reaching 'No Child Left Behind' (NCLB) educational initiative has resulted in a dramatic rise in the use of standardized tests intended to provide quantitative evidence of student progress (Ravitz, 2010). Since English language learners (ELLs) must be included in these assessments, and assessments are often administered only in English, they have the potential to immediately disadvantage many ELLs (Ravitz, 2010; Menken, 2009). These are high stake assessments for these students since, as Menken (2009, p. 205) notes, "ELLs are disproportionately failing high-stakes tests and being placed into low-track remedial education programs, denied grade promotion, retained in grade, and/or leaving school."

These developments within the North American context have motivated some to argue for a re-evaluation of test developer responsibilities and much greater recognition of the social impact that language tests have on school-age L2 learners (e.g., Fox and Cheng 2007; Chalhoub-Deville, 2009). In the Asian context, increasing concern has been expressed about the potential impact of English language tests on the educational context of learners of all ages (Cheng, 2005, 2008). Ross (2008) has stated that English language testing is one of the "main pillars of the school gate [that] is used to control access to selective middle schools and high schools" (p.7). Within such test-intensive environments, as Rea Dickens and Scott (2007) point out, language testing practices have the "potential for considerable impact in terms of affective factors" (p. 3).

Assessment practices and adolescent motivation

While still under-researched, a number of studies have investigated the relationship between assessment practices and the motivation to learn (Black and Wiliam, 1998; Crooks, 1988; Harlen and Deakin Crick, 2003) and a number of these studies have discovered a unique relationship for adolescent students. For example, in a study investigating the impact of classroom assessment practice on the attitudes, comprehension, and attributions of 304 public school students (spanning grades 4 to 11), Evans and Engelberg (1988) discovered that simple grading concepts were better understood by those aged 11 or older. However, lower-achieving students were more likely to believe that "getting good grades is something beyond their control or influence" making it "unlikely that they will work at doing better" (p. 51).

Exploring the relationship between standardized testing and motivation more specifically, Paris et al. (1991) investigated almost 1000 students' perceptions (grades 2 to 11) of standardized achievement tests and found that by adolescence "less successful students in particular feel powerless to control their own success in school and may feel victimized by tests that confirm their own low performance" (p. 14). Similarly, Pollard et al. (2000) conducted a large-scale study of the impact of national standards-driven educational reform in England and Wales and found that adolescent students (compared to younger learners) were more likely to attribute the results of their tests to internal factors (e.g. effort or ability) rather than external factors (e.g., practice). The authors report that the increase in

standardized achievement testing led to greater differentiation among students (by both teachers, peers, and the students themselves) and that "by the end of the SATs, children seemed more sure than they had ever previously been of who was 'bright' and who was 'thick'" (p. 238). These studies support the assertion that among adolescent learners, low achievers are more likely to become "overwhelmed by assessments and demotivated by constant evidence of their low achievement thus further increasing the gap" (Harlen and Deakin Crick, 2003, p. 196).

Studies explicitly investigating the relationship between language testing experience and second language motivation among young adolescent language learners (or test-takers) are quite scarce. There is a significant body of research exploring the impact of language testing on teachers and classrooms (e.g., Chapman and Snyder, 2000; Cheng, 2005; Qi, 2007) and the socio-political systems surrounding them (e.g., McNamara and Roever, 2006; Shohamy, 2001); however, impact on the "ultimate stakeholder", the individual language learner, has received far less attention (Cheng et al., 2010, p. 222). Additionally, although there is an impressive body of research focused on young adolescent learners in the literature on language teaching and learning (e.g., Guilloteaux and Dörnyei, 2008; Roessingh et al. 2005; Swain and Lapkin, 1995), very few studies have focused on these learners in relation to their language assessment environment (Fox and Cheng 2007), as older adolescents, adults, and young English language learners (under 12 years) have tended to occupy researcher interest.

The present study: language testing and L2 motivation

The present study aims to contribute to a deeper understanding of the complex relationship that exists between language assessment practices and the motivation to learn English among young adolescent language learners. Undoubtedly, there are a myriad of factors that can potentially impact a students' motivation to learn a language; however, we feel the role that high stakes language testing experience plays in this process needs much more consideration, especially within highly test-intensive environments (for additional discussion, see Haggerty and Fox: Test intensity, language testing experience, and the motivation to learn English in South Korea, forthcoming).[1]

In this study, we explore the relationship between L2 testing experience and L2 motivation amongst a group of young adolescent students (aged 12–15) enrolled in a private language school (or *hakwan*) in Seoul, South Korea ($N = 341$). We investigated L2 testing experience by considering both the time students spent preparing for language tests as well as their experiences taking either a middle-school or university level high-stakes standardized language test. To explore the relationship to L2 motivation, we draw on constructs from a theoretical model (the L2 Motivational Self System) which has been developed by Zoltan Dörnyei and his colleagues for well over a decade (Dörnyei and Ushioda, 2009).

This research has also been inspired by a unified conception of test validity that includes the social consequences of test use (Messick, 1989, 1995), the need to seek evidence from the social environment as part of an ongoing validation process (Koch and Deluca, 2012; McNamara and Roever, 2006), and the need for greater incorporation of test-takers' perspectives in the validation of large-scale English language tests (Cheng and DeLuca, 2011).

To situate the present study, the section below provides additional background information about the testing context for young adolescents in South Korea (hereafter Korea

or Korean) and the theoretical model from which the L2 motivation constructs used in this study were derived.

English language testing practices in South Korea

This research was conducted within the Korean educational context, one that has been characterized as highly test-intensive (Choi, 2008) in which many language teachers tend to adopt a 'teach to the test' mentality (Hwang, 2003; Koh, 2007; Li, 1998) and many students feel compelled to attend private language schools (or *hakwans*) to improve their English. A *hakwan* is as a type of private tutoring which is conducted in a "private for-profit learning institution" with instruction given in a "classroom-like setting" (Kim and Lee, 2010, p. 261). The proliferation of hakwans has been a highly contentious educational issue in Korea for many decades. It has been estimated that about 75 % of primary and secondary students receive some kind of private tutoring, requiring families to spend an amount "equivalent to about 80 % of governmental expenditure on public education" (ibid.). Korean governmental responses to private tutoring since the 1950s have included failed attempts to ban the practice, greater regulation of institutions and teachers (e.g., hours of operation, teaching certification), and various public school initiatives (e.g., after-school programs) (Kim and Chang, 2010). Despite these efforts, the pervasive and powerful influence of hakwans on the Korean education system has continued largely unabated.

Korean middle-school students are required or encouraged to take a wide variety of language tests. Two well-established publicly available language tests in this context are the Test of English as a Foreign Language (TOEFL), including the paper-based (PBT), computer-based (CBT), and internet-based (iBT) versions, and the Test of English Proficiency (TEPS), developed by Seoul National University in South Korea (Choi, 2008). While primarily serving gatekeeping functions for universities and companies, these tests have also been used to inform the admission decisions of a number of private and public high schools. This practice has led to increasing pressure for many young learners to prepare for and take one of these tests before the end of middle school.

It is difficult to determine precisely how many Korean middle-school students have taken such standardized tests over the last decade, but there are indications that it is a large number. The Korean Ministry of Education, Science and Technology has stated that 39,700 South Korean students between the ages of 13 and 15 registered to take the TOEFL or TOEIC in 2007 alone (Kang, 2010). In August, 2011, Korean media announced that a 12-year old student had achieved a perfect score on the TOEFL iBT, eclipsing the feat previously accomplished by a 13-year old student (Kim, 2011). For young adolescents preparing for English language tests in Korea, the message being sent is clear: aim high, do it early, and be perfect. This, as Choi (2008) has cautioned, creates "unwarranted pressure" for young learners to take language tests beyond their cognitive capacity, produces a "mismatch between the difficulty of test input and the ability of test takers [that] leads to invalid consequences", and raises "ethical issues regarding children's right to learn in an appropriate manner" (p. 53).

Given the nature of these assessment practices, we concur with Elwood's (2013) conclusion that it is critical to "consider what an ethical assessment environment or milieu might look like for young people who are caught up at the centre of so much assessment practice" (p. 207). To move in this direction in the field of second language

assessment, it is necessary to better understand the impact of current assessment practices on young adolescents' L2 motivation, and investigate potential consequences of test use and testing experience, particularly when high-stakes tests are involved. Just what is at stake when the bar is raised so high for young Korean adolescent L2 learners? What is the impact of administering standardized language proficiency tests such as the TOEFL iBT (which measures university-level language) to 11 and 12-year olds? Is there any evidence of impact on their motivation to learn English?

The motivation to learn a second language

Motivation has long been recognized as an important factor in L2 learning (Dörnyei, 2001). For the last decade, Dörnyei and his colleagues (e.g., Csizér and Dörnyei, 2005; Dörnyei and Ushioda, 2009) have been investigating L2 motivation within international contexts. This work has led to the development of a theoretical model termed the *L2 Motivational Self System* (Dörnyei, 2009) which consists of three main components: 1) the Ideal-L2 Self; 2) the Ought-to L2 Self; and 3) the L2 learning experience. A brief explanation of these three components is necessary before presenting the L2 motivation constructs that guided this study.

According to Dörnyei (2009) the *Ideal L2 Self* is composed of idealized images we hold about ourselves as future L2 users. The *Ought-to L2 Self* is composed of idealized images generally intended to please others or avoid negative repercussions. These two components are theoretically informed by future-selves research (within social psychology) which posits that our imagined "possible selves" act as "future self-guides" and, as such, are experienced in the here and now as "self-states" (Dörnyei, 2009, p. 16). It is further argued by Dörnyei and his colleagues that the Ideal L2 Self is very influential in shaping motivational behaviour. Unlike the Ought-to L2 Self, it is amenable to motivational strategies (or interventions) designed to stimulate and enhance this vision (p. 32). The third component, the L2 Learning Experience, involves "situated, 'executive' motives related to the immediate learning environment and experience (e.g. the impact of the teacher, the curriculum, the peer group, the experience of success)" (p. 29).

While drawing inspiration from Dörnyei's (2009) L2 Motivational Self System, the L2 motivation constructs utilized for the present study were primarily based on a large-scale structural equation modelling (SEM) study conducted by Taguchi, Magid, and Papi (2009), which was designed to assess the suitability of the model in three Asian countries (China, Japan, and Iran). Their results provided validation evidence to support the model and helped them to understand "certain cross-cultural differences" (Taguchi et al., 2009, p. 88). In their final SEM model, they identified eight factors as significant contributors to L2 motivation in these three language learning contexts. These formed the basis for eight of the ten constructs used in this study. Two additional constructs were incorporated from the results of a pilot study (Haggerty, 2011b) conducted in South Korea with the population of test takers of interest in the present study. All ten constructs (along with a definition for each) are listed in Table 1.

While most of the definitions for these constructs are fairly self-explanatory, the 'promotion' and 'prevention' aspects of instrumentality require some further elaboration. In the L2 motivation literature, instrumentality generally refers to language learners' perceptions about the benefits of attaining L2 proficiency (Csizér and Dörnyei, 2005, p. 21). These are often based on personal and practical objectives such as educational/career

Table 1 L2 motivation constructs

Construct	Description
1. Ideal L2 Self[a]	Idealized images about being a successful L2 user. A strong vision can be a powerful motivator to harmonize our actual and future self
2. Ought-to L2 Self[a]	Idealized images about one's future L2 self. Influenced by a sense of obligation to meet others' expectations or please them
3. Instrumentality (promotion)[a]	The regulation of personal goals to become more successful. (e.g., attaining high proficiency in English to improve career potential)
4. Instrumentality (prevention)[a]	The regulation of duties and obligations in order to avoid negative outcomes (e.g., studying English in order to pass an important exam)
5. Attitudes to L2 Learning[a]	Situation-specific motives related to the immediate learning environment and experience (e.g., impact of peers, teachers, curriculum, etc.)
6. Attitudes to L2 Culture and Community[a]	The learner's attitudes toward the target language community including interest in its cultural products (e.g., movies, music, etc.)
7. Family Influence[a]	The influence of (active or passive) family members on the language learning activities engaged in by the learner
8. Intended Effort[a]	The degree of effort a language learner intends to exert towards language learning activities
9. Attitudes to L2 Testing[b]	A language learners' current disposition towards L2 testing (comprised of value, accuracy, impact, usefulness, fairness, importance, predictability, and trust of L2 tests or L2 test results)
10. L2 Social Value[b]	Student's perception of the importance of an L2 within their own sociocultural context

[a]adapted from Dörnyei (2009) and Taguchi et al. (2009)
[b]adapted from Haggerty (2011b)

advancement or a desire to travel. However, Dörnyei (2009) asserts that from a self-perspective, instrumentality can be further divided into a *promotion* focus related to our "hopes, aspirations, advancements, growths and accomplishments" and a *prevention* focus related to "safety, responsibilities and obligations" (p. 28). The Taguchi et al. (2009) study provided further support for the division of this construct into these two separate dimensions.

These ten L2 motivation constructs guided our exploration of the relationship between testing experience and the motivation to learn English among young adolescent language learners in Korea. Instead of using SEM (which seeks causal relationships), this study utilizes exploratory factor analysis (EFA) to identify the most salient L2 motivation factors in this research context. Specifically, this study was guided by the following research questions:

1. Which L2 motivation constructs are most salient among the young Korean adolescents in this study?
2. What relationships exist between the L2 motivation constructs identified and: a) participants' gender and grade level, b) the amount of time participants prepare for English language tests, and c) whether participants have taken a university-entrance level language test (the TOEFL or TEPS)?

Method

The research site

This study was conducted within a private language school also referred to as an 'academy,' 'cram school' or 'hakwan'. Located in the capital of Seoul, Korea, this school specialized in improving middle-school students' overall English language skills as well as preparing them for a number of standardized language tests. An emphasis on test preparation is

typical of such schools (Kim and Chang, 2010; Kim and Lee, 2010). We originally contacted a number of other hakwans in various locations around Seoul, but only one agreed to participate in the study. We had hoped to include short answer questions on the questionnaire as well as conduct follow-up interviews. However, we were told by our designated contact (one of the more experienced teachers at the school) that this would be too distracting for their students who needed to focus on their English studies. However, we were welcome to administer a questionnaire composed of only closed-ended questions. The research site was visited on numerous days over a two week period in order to administer the questionnaire in all classes. Our local contact assisted in explaining the research to students in their first language and answering any questions they had.

Participants

The participants for this study were all middle-school students ($N = 341$). A total of 395 students were registered at the time of the study; however, 49 students were absent and 5 students elected not to participate (86 % response rate). A total of 141 (41.3 %) were male and 200 (58.7 %) were female. The ages of the participants ranged from 12 to 15 with a mean age of 14.3 years. 59 (17 %) of participants were in their first year of middle school, 60 (18 %) in their second, and 222 (65 %) in their third and final year. Of those who reported the amount of hours spent each week preparing for L2 tests ($n = 319$), 81 (25 %) indicated 'zero hours', 90 (28 %) 'one to three hours', 78 (25 %) 'four to six hours', and 70 (22 %) 'seven hours or more'. Finally, 229 (67 %) of the participants reported not taking a university-entrance level L2 test (TOEFL or TEPS) while 112 (33 %) reported they had. When separated by age, 3/15 (20 %) of 12-year olds, 13/56 (23 %) of 13-year olds, 34/98 (35 %) of 14-year olds, and 62/172 (36 %) of 15-year-olds reported taking one of these tests.

Instrument

A questionnaire[2] was developed to elicit participants' strength of agreement based on a 6-point Likert scale (1 = strongly disagree to 6 = strongly agree, for statements; 1 = not at all to 6 = very much, for questions). All questionnaire items were provided in Korean and English. To assess the suitability of translated questions, we followed a team-based approach involving two native-Korean teachers and two native-Canadian teachers. The main focus was to maximize intelligibility for middle-school students. In the end, 28 items were included from Taguchi et al. (2009) (see Additional file 1: Appendix A for the English wording of all questionnaire items). An additional twelve items (see Additional file 2: Appendix B) were drawn from the pilot study (Haggerty 2011b). To check for consistency in participant responses, two items were negatively worded. Originally we had hoped to include open-ended questions and conduct interviews with some students after questionnaire responses were analyzed; however, we were not given permission to do so at this institution. We were told that this would unnecessarily distract students from their studies. Given the difficulty we experienced gaining access to other privately-run institutions, we decided to collect as much data as we could without interfering too much in the day-to-day operation of the school.

Data analysis

All data were analyzed using a number of statistical procedures available within the Social Sciences Statistical Package (SPSS v.16). The threshold significance level for all tests was established at $p < .05$. We followed a 3-step data analysis procedure:

1. We identified the most salient L2 motivation factors for these participants following established EFA procedures (see below).
2. After L2 motivation factors were identified, we calculated factor scores for all participants in order to make subgroup comparisons.
3. We conducted analysis of variance (ANOVA) for variables with three or more subgroups and Pearson product–moment correlation for those with only two.

Some explanation of exploratory factor analysis (EFA) might be helpful before describing some of the other statistical techniques we utilized. EFA is a quantitative method that enables researchers to investigate the co-relationships amongst numerous variables to determine if they can be reduced and summarized in a "smaller number of latent constructs" (Thompson, 2004, p. 10). When utilizing EFA, a researcher may have some expectations about the underlying constructs involved, but it does not require them to be firmly identified prior to conducting the statistical procedure (Thompson, 2004, p. 5). On the other hand, confirmatory factor analysis (CFA) requires that constructs be modelled prior to conducting any statistical analysis (as was done in Taguchi et al., 2009). The success of the model is then measured by various goodness of fit indices. Since these constructs had not been tested with young adolescent learners in this specific context, we chose EFA over CFA as it better suited the purposes of this exploratory study.

Before proceeding with factor analysis, it is important to consider whether a specific data set is conducive to factor analytic techniques. To better determine this, three measures were considered in this study. First, in terms of sample size, Gorsuch (1997) has suggested a minimum ratio of five participants per variable and no less than 100 participants. For this study, a modest sample to item ratio of 8.5:1 (341 participants, 40 questionnaire items) was achieved (MacCallum et al., 1999). In addition, the Kaiser-Meyer-Olkin (KMO), a measure of sampling adequacy, was .896 which has been characterized as "great" (Field, 2009, p. 647). Finally, Bartlett's test of sphericity, a measure to determine whether there is sufficient variability among items, was significant ($p < .05$).

In terms of the EFA procedure followed, Principal axis factoring with a Promax rotation was used since we expected constructs to correlate and it offered the most clearly interpretable results. To guide factor selection, two standards were set: 1) they obtained eigenvalues of 1 or more (the Kaiser criterion), and 2) they were located at or above a point of inflexion (or 'elbow') on a scree plot. A data reduction procedure removed weakly-loaded and cross-loaded items (< .30). For additional information about these procedures see Field (2009, p. 639) and for additional details about these results see Haggerty (2011a).

After the final factor solution was identified, factor scores for all participants were calculated. Factor score distributions were inspected to ensure that they satisfied the normality assumptions of the statistical tests to be used (DiStefano et al. 2009). For learner-background variables with three or more categories (Grade Level and L2 Test Preparation Time), separate one-way ANOVA tests were conducted for each L2 motivation factor. For dichotomous variables (Gender[3] and University-level L2 Test Experience), Pearson point-biserial correlations were calculated. Levene's test for homogeneity of variance indicated four of the five factors satisfied this assumption. In these cases, Tukey's HSD post hoc test was used. For the factor 'Ideal L2 Self', homogeneity of variance could not be assumed, but a Welch's F test indicated subgroup means were statistically significant ($p < .05$). For this factor, the Games-Howell post hoc test was used.

Results and discussion

Research question #1: Salient L2 motivation factors

Our initial research question sought to identify which L2 motivation constructs were most salient. Figure 1 is a visual representation of the five L2 motivational factors identified and their relationship to the original constructs. The alpha level achieved for all five factors was above .70 (four were above .80). The first factor identified, Ought-to L2 Self + Instrumentality (prevention) was based on a preponderance of factor loadings from two theoretical constructs. The remaining four factors were each identified based on a preponderance of factor loadings from one theoretical construct. The factor 'Ought-to L2 Self + Instrumentality (prevention)' accounted for the largest proportion of the variance (27.7 %). This factor, being a conflation of two theoretical constructs, incorporated a number of items revolving around the need to avoid shame and embarrassment, a belief that studying the L2 is necessary, and the desire to gain approval from significant others. This conflation can be explained based on the results of previous studies that have found 'Ought-to L2 Self' to be more closely associated with 'Instrumentality (prevention)' and 'Ideal L2 Self' to be more closely associated with

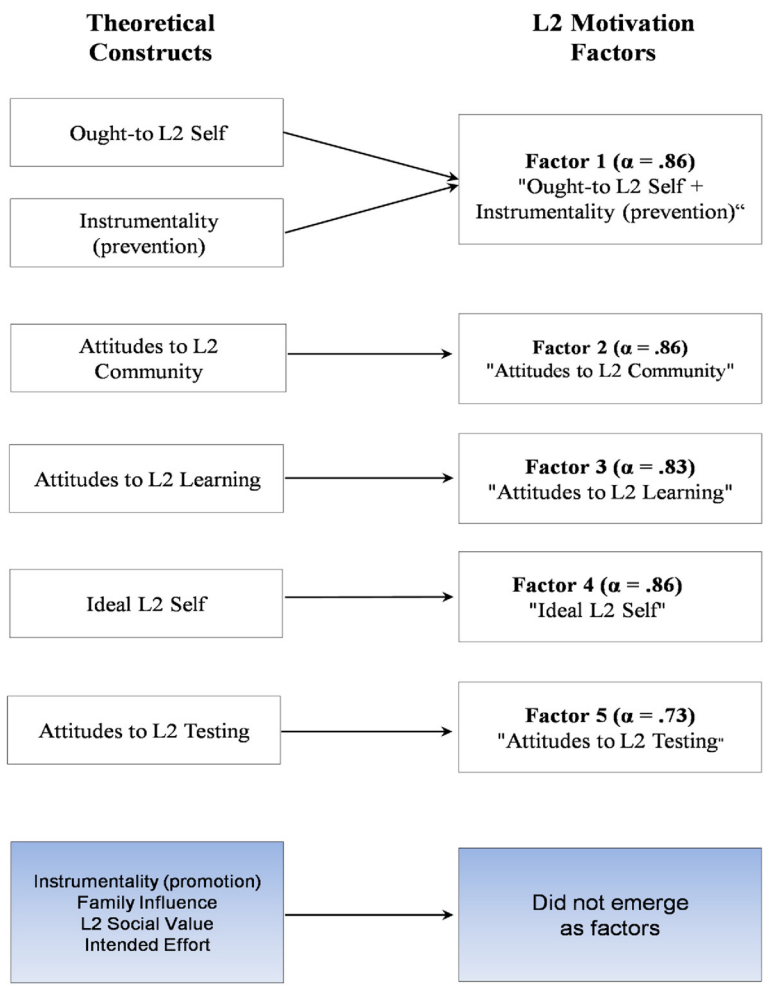

Fig. 1 Theoretical concepts and L2 motivation factors identified

'Instrumentality (promotion)' (Dörnyei, 2009, p. 31). Since 'Instrumentality (promotion)' was one of four constructs that did not emerge as factors in this study, we were unable to explore its relationship to 'Ideal L2 Self' in more detail.

Research question #2a: Gender, Grade Level, and L2 motivation

We found no statistically significant differences between Gender or Grade Level for all five L2 motivation factors ($p > .05$). However, female participants (as a group) did hold slightly more positive 'Attitudes to L2 Community' than male participants ($r = .11$, $p = .048$). Given these results, it was decided to treat the sample as one homogenous group (young adolescent learners) for the subsequent analysis of their L2 testing experience. Future studies will hopefully shed more light on the complex motivational changes that may occur as students progress through middle school (i.e., as language test results become increasingly important).

Research question #2b: L2 Test Preparation Time and L2 motivation

We compared the amount of time participants reported preparing for L2 tests with each of the five L2 motivation factors. Separate one-way ANOVAs were performed for L2 Test Preparation Time and each L2 motivation factor. The results for each test are listed in Table 2.

As indicated in the table, there was a significant association ($p < .05$) between the amount of time spent preparing for L2 tests and each of the five L2 motivation factors. Post hoc analyses revealed significant differences between groups who reported '0 h per week' and '7 h or more per week' for all five motivation factors. There were also significant differences discovered for groups who reported '0 h per week' and '4-6 h per week' for 'Attitudes to L2 Learning' and 'Attitudes to L2 Testing' and between groups who

Table 2 One-way ANOVA for L2 test preparation and L2 motivation

L2 Motivation Factors		*Sums of Squares*	*df*	*Mean Square*	*F*	*Sig.*	*η2*
Ought-to L2 Self + Instrumentality (prevention)	Between Groups	14.79	3	4.93	5.55	.001	.05
	Within Groups	271.61	316	.89			
	Total	286.40	319				
Attitudes to L2 Community	Between Groups	17.23	3	5.74	6.55	.000	.06
	Within Groups	268.27	316	.88			
	Total	285.50	319				
Attitudes to L2 Learning	Between Groups	22.72	3	7.57	8.98	.000	.08
	Within Groups	258.03	316	.84			
	Total	280.74	319				
Ideal L2 Self[a]	Between Groups	24.71	3	8.24	9.73	.000	.09
	Within Groups	259.16	316	.85			
	Total	283.87	319				
Attitudes to L2 Testing	Between Groups	29.94	3	9.98	14.07	.000	.12
	Within Groups	217.02	316	.71			
	Total	246.96	319				

[a]Homogeneity of variance could not be assumed, Welch's F was significant $p < .001$, Games-Howell post hoc used

reported '1-3 h per week' and '7 h or more' for 'Ideal L2 Self'. The effect sizes for L2 test preparation time and each L2 motivation factor are generally in the medium range ($\eta 2 = .0588$ to .1379) (Cohen, in Richardson, 2011).

Overall, the ANOVA results suggest that L2 motivation was significantly associated with the amount of time spent in L2 test preparation; those spending less time being generally less motivated than those spending more. These results could be explained by referring to a number of expectancy-value theories of motivation (see Dörnyei, 2001, for an extensive review). For example, it could be argued that students who prepared more for L2 tests may simply have valued the activity more, anticipated a greater chance of success, and were thus more highly motivated. However, this simplistic interpretation would fail to shed much light on how language learner thought and behaviour is influenced by the wider sociocultural environment impinging on all language learners, and how motivational behaviour is shaped over time and space, a longitudinal process captured more accurately by the notion of "investment" (see Darvin and Norton, 2015; Norton, 2013).

Undoubtedly, there are many pieces to the complex L2 motivational process at work in this educational context. Very likely participants' family influences, previous success in language learning, pedagogical practices, peer pressure, and other sociocultural factors have sent important signals to these participants. However, we would argue that the test-intensive environment itself, largely downplayed or ignored in previous motivation research, is a salient factor requiring much deeper investigation. To help redress this unfortunate oversight, we have also investigated the relationship between the completion of a university-entrance level L2 test and young adolescent L2 motivation.

Question #2c: University-level L2 Test Experience and L2 motivation

To address this research question, we compared university-entrance level testing experience (operationalized as those who have taken or have not taken either the TOEFL or TEPS) with each of the five L2 motivation factors. Pearson point-biserial correlation coefficients were calculated and the results are listed in Table 3. Significant correlations ($p < .05$) were discovered for four of five factors: 'Attitudes to L2 Community' ($r = .17$), 'Attitudes to L2 Learning' ($r = .21$), 'Ideal L2 Self' ($r = .17$), and 'Attitudes to L2 Testing ($r = .20$). There was no significant difference found between the subgroups for the factor 'Ought-to L2 Self + Instrumentality (prevention)' ($r = .03$).

The results indicate that the middle-school students in this study who had not taken one of the university-entrance level English proficiency tests available to them (the TOEFL or TEPS) also tended to have more negative L2 motivation. Although the effect sizes are between the small (.10) to medium (.30) range (Cohen, 1992), given the complexity involved in studying L2 motivation, we would argue that these results are an indication that completing a more advanced language proficiency test (TOEFL or TEPS) is strongly associated with participants' overall level of L2 motivation. Interestingly, the

Table 3 Point-biserial correlations for TOEFL/TEPS and L2 Motivation

TOEFL/TEPS		Ought-to L2 Self + Instrumentality (prevention)	Attitudes to L2 Community	Attitudes to L2 Testing	Ideal L2 Self	Attitudes to L2 Learning
(no/yes)	Correlation	.03	.17[a]	.21[b]	.17[a]	.20[b]
	Sig (2-tailed)	.62	.00	.00	.00	.00

[a]Correlation is significant at the .01 level (2-tailed)
[b]Correlation is significant at the .001 level (2-tailed)

factor 'Ought-to L2 Self + Instrumentality (prevention)' was not significantly associated with the completion of the TOEFL or TEPS. This suggests that these middle-school students felt similar pressure to study English (to please others and avoid the results of failure) regardless of whether they had taken one of these high-level tests.

The results of this research suggest that high-stakes L2 testing practices can play a significant role in mediating young adolescent L2 motivation. It also supports the contention that large-scale high-stakes language tests can significantly impact young adolescents' L2 motivation in highly test-intensive environments like the one considered here. According to informants working within this context, middle-school students who have not demonstrated an ability to take a high-level test (based on their performance on simulated tests) are generally placed in lower level classes. Only students who demonstrate a fairly high level of proficiency are generally encouraged to take a university-level entrance test. These kinds of placement decisions, largely (mis) informed by the ability of middle-school students to complete proficiency tests well beyond their cognitive level, may have a lasting impact on students' learning trajectories. As Roderick and Engel (2001) have noted, "students need to believe that they can achieve the goal given their skills" (p. 200). Given what many other researchers have discovered about the relationship between testing and motivation among adolescent learners (Black and Wiliam, 1998; Crooks, 1988; Harlen and Deakin Crick, 2003), and the cognitive level of the language tests influencing this context, serious issues emerge concerning the effectiveness, ethicality, and validity of this testing practice.

Conclusion

Our findings suggest that the amount of time middle-school students spend on L2 test preparation and their ability or inability to take a university-entrance level language test (TOEFL iBT or TEPS) may play an important role in mediating L2 motivation in this test-intensive context. Those spending less time preparing for L2 tests, or those who have not taken a university-entrance level L2 test, tended to have more negative L2 motivation. These findings support the assertion that the language assessment environment may be adversely affecting the motivation to learn English for a significant number of language learners in this context. It is important to note that while 33 % of the participants were given the opportunity to take a university-level language proficiency test, 67 % were not. The score obtained on this test may not be as important as getting the opportunity, a signal of achievement in and of itself. As mentioned earlier, students are generally not encouraged or permitted to take such tests unless they can perform fairly well on simulated tests first. It is also important to note that while participants differed significantly on four of the five measures of L2 motivation based on whether they took a TOEFL or TEPS, this variable did not correlate significantly with 'Ought-to Self + Instrumentality (prevention)', suggesting many students shared a similar sense of obligation and desire to please others in order to avoid the negative consequences of failure.

Overall, these results support the conclusion that setting the "proficiency test" bar too high may create unrealistic expectations for many and, in turn, de-motivate students unable to reach these high standards, an assertion supported by a number of studies that have investigated the relationship between testing and motivation among adolescent learners (Black and Wiliam, 1998; Crooks, 1988; Harlen and Deakin Crick, 2003). The pressure to take a language test that is beyond the cognitive capacity of

most young adolescent students, it can be argued, is likely to have an adverse effect on the L2 motivation of young adolescent learners who, while under the same pressure to perform well on high-level standardized language tests, are simply unable to do so. The "unwarranted pressure" created by these testing practices (Choi, 2008) may have appreciable long-term effects on young adolescent L2 motivation; an issue in need of much further investigation, particularly in test-intensive educational contexts like the one examined here.

In considering these findings, it is important to acknowledge the limitations of this study. In terms of the EFA procedure followed for this investigation, the results were limited by the number of items included per construct and the sample size. While we are satisfied with the balance achieved given the purposes of this exploratory study, future studies should attempt to increase the number of items included for each construct under inquiry (at least five). Also, while the sample size for this study was adequate for EFA, a larger sample size would increase the potential for meaningful factors to emerge. Also, since the sample consisted of middle-school students in one private language school, these results cannot be generalized to the wider population.

In this study, we were limited to the collection of quantifiable questionnaire data and thus unable to triangulate our findings through the collection of qualitative data (e.g., written responses or one-on-one interviews), something we feel is essential for getting a richer and more meaningful understanding of highly complex sociocultural and sociocognitive processes. In addition, this questionnaire was administered only once and thus we were unable to track changes that might be occurring over time. Additional research that can better overcome these limitations is definitely needed, particularly in test-intensive EFL contexts involving young adolescent language learners.

Finally, it is important to note that this study did not directly probe the impact of L2 testing practices on language learners' desire or willingness to commit to, or invest in, language learning over time (Darvin and Norton, 2015; Norton, 2013). Despite these limitations, this research revealed significant patterns in the relationship between young adolescents' L2 motivation and their L2 testing experience; findings we hope will spur greater interest and investigation.

This study affirms the value of continually seeking evidence from the social context in which high-stakes language tests are in use in order to better inform ongoing validity arguments (e.g., Fox 2003; McNamara and Roever, 2006; Messick, 1989, 1995). More specifically, these results suggest a need to make more informed decisions about what level and type of testing is appropriate for specific age groups. Although in some settings students can "challenge" a course offered at a higher level, such challenges do not typically occur across exceedingly different developmental levels (i.e., middle school to university). While challenging young adolescent learners is certainly an important part of the educational process, what level of challenge is appropriate and at what age? If testing practices such as those reported in this study can be shown to be detrimental to young adolescents, who is (or should be) responsible for addressing the issue? There is also an important socio-ethical issue associated with the economic burden such an extreme emphasis on English puts on families who cannot afford to send their children to the "best" English schools (Spolsky, 2007).

These results support the contention that there is a benefit, if not a necessity, for test developers and other stakeholders to acknowledge the wider social context in which

tests are deployed. This is especially urgent when there are indications, such as those suggested by this study, that L2 testing practices may be contributing to an educational culture among young adolescent language learners that motivates some learners at the expense of many others. If one of our goals as parents, educators, curriculum developers, and hopefully as test developers and administrators, is to develop an environment conducive to learning, it is necessary to better understand not only what test takers are doing on a test, but also what tests might be doing to them.

In our view, high-stakes test developers should continue to be encouraged to move beyond the 'what and how' of their tests as much as possible. Such efforts could include, amongst others: 1) statements about the age-appropriateness of a language test; 2) statements about the potentially detrimental sociocognitive and sociocultural effects involved in test misuse; and 3) a mechanism for collecting information about the social consequences of test deployment as part of an ongoing test validation process. In addition, to assist test developers, it is hoped that future studies will continue to shed more light on the social consequences (both sociocognitive and sociocultural) of L2 assessment practices on language learners around the world.

Endnotes

[1]These results are also reported in a chapter in a forthcoming book entitled, *Current Trends in Language Testing in the Pacific Rim and the Middle East: Policies, Analyses and Diagnoses*, edited by Vahid Aryadoust and Janna fox, Cambridge Scholars Publishing. That chapter reports on some results not presented here and discusses additional contextual issues surrounding language testing practices (e.g., outcomes-based curriculum and test intensity).

[2]In the interest of brevity, the questionnaire is not included here. However, it is available on request by emailing the corresponding author.

[3]We acknowledge and encourage the continual problematization of this binary socially-constructed notion. However, for the purposes of this study we felt it worthwhile to explore this variable given its history in the literature.

Additional files

Additional file 1: Appendix A: Questionnaire Items for L2 Motivation Constructs (Taguchi et al. 2009).
Additional file 2: Appendix B: Questionnaire Items for L2 Motivation Constructs (Haggerty 2011b).

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

Authors' contributions
Both authors read and approved the final manuscript.

Author details
[1]Department of Language and Literacy Education, University of British Columbia, Vancouver, BC, Canada. [2]School of Linguistics and Language Studies, Carleton University, Ottawa, ON, Canada.

Received: 8 April 2015 Accepted: 30 June 2015
Published online: 26 July 2015

References

Black, P, & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*, 7–74.
Chalhoub-Deville, M. (2009). The intersection of test impact, validation, and educational reform policy. *Annual Review of Applied Linguistics, 29*, 118–131.

Chapman, D, & Snyder, CW. (2000). Can high-stakes national testing improve instruction: reexamining conventional wisdom. *International Journal of Educational Development, 20*, 457–474.
Cheng, L. (2005). *Changing language teaching through language testing: a washback study*. Cambridge: Cambridge University Press.
Cheng, L. (2008). Washback, impact and consequences. In E Shohamy & N Hornberger (Eds.), *Encyclopedia of language and education* (Vol. 7, pp. 349–364). New York: Springer.
Cheng, L, & DeLuca, C. (2011). Voices from test-takers: further evidence for language assessment validation and use. *Educational Assessment, 16*, 104–122.
Cheng, L, Andrews, S, & Yu, Y. (2010). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing, 28*, 221–249.
Choi, I. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing, 25*, 39–61.
Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.
Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*, 438–481.
Csizér, K, & Dörnyei, Z. (2005). The internal structure of learning motivation and its relationship with language choice and learning effort. *The Modern Language Journal, 89*, 19–36.
Darvin, R, & Norton, B. (2015). Identity and a model of investment in applied linguistics. *Annual Review of Applied Linguistics, 35*, 36–56.
DiStefano, C, Zhu, M, & Mindrilla, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation, 14*, 1–11.
Dörnyei, Z. (2001). *Teaching and researching motivation*. Essex, England: Pearson Education Ltd.
Dörnyei, Z. (2009). The L2 motivational self system. In Z Dörnyei & E Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Bristol: Multilingual Matters.
Dörnyei, Z, & Ushioda, E. (Eds.). (2009). *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters.
Elwood, J. (2013). Educational assessment policy and practice: A matter of ethics. *Assessment in Education: Principles, Policies, and Practice, 20*, 205–220.
Evans, E, & Engelberg, R. (1988). Pupils' perceptions of school grading. *Journal of Research and Development in Education, 21*, 44–54.
Fairbairn, S, & Fox, J. (2009). Inclusive achievement testing for linguistically and culturally diverse test takers: Essential considerations for test developers and decision makers. *Educational Measurement: Issues and Practice, 28*, 10–24.
Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.
Fox, J. (2003). From products to process: An ecological approach to bias detection. *International Journal of Testing, 3*, 21–48.
Fox, J, & Cheng, L. (2007). Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first (L1) and second (L2) language test takers. *Assessment in Education, 14*, 9–26.
Gorsuch, R. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment, 68*, 532–560.
Guilloteaux, M, & Dörnyei, Z. (2008). Motivating language learners: A classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL Quarterly, 42*, 55–77.
Haggerty J. (2011a). *High-stakes language testing and young English language learners: Exploring L2 motivation and L2 test validity in South Korea (Master's thesis)*. Retrieved from ProQuest Dissertations and Thesis. (AAT 881645773).
Haggerty, J. (2011b). An analysis of L2 motivation, test validity and Language Proficiency Identity (LPID): A Vygotskian perspective. *Asian EFL Journal, 13*, 198–227.
Harlen, W, & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education, 10*, 169–207.
Hwang, HJ. (2003). *The impact of high-stakes exams on teachers and students*. Montreal, Canada: Unpublished MA dissertation, McGill University. Retrieved from ProQuest Dissertations and Theses. (AAT 305243783).
Kang, S. (2010, July 15). TOEFL for juniors to debut here. Korea Times. Retrieved July 20th, 2010 from http://www.koreatimes.co.kr
Kim, T. (2011, August 26). 12 year old girl gets perfect TOEFL score. Korea Times. Retrieved August 28th, 2011 from http://www.koreatimes.co.kr
Kim, J, & Chang, J. (2010). Do governmental regulations for cram schools decrease the number of hours students spend on private tutoring? *KEDI Journal of Educational Policy, 7*, 3–21.
Kim, S, & Lee, J. (2010). Private tutoring and demand for education in South Korea. *Economic Development and Cultural Change, 58*, 259–296.
Koch, M, & DeLuca, C. (2012). Rethinking validation in complex high-stakes assessment contexts. *Assessment in Education, 19*, 99–116.
Koh, YK. (2007). *Understanding Korean education; Korean education series: School curriculum in Korea (Vol. 1). Korean Educational Development Insitute (KEDI)*. Retrieved March 5, 2010 from http://eng.kedi.re.kr.
Li, D. (1998). "It's always more difficult than you plan or imagine": Teachers' perceived difficulties in introducing the communicative approach in South Korea. *TESOL Quarterly, 32*, 677–703.
MacCallum, R, Widaman, K, Zhang, S, & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84–89.
McNamara, T, & Roever, C. (2006). *The social dimension of language testing*. Malden, MA: Blackwell.
Menken, K. (2009). No child left behind and its effect on language policy. *Annual Review of Applied Linguistics, 29*, 103–117.
Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Education Researcher, 18*, 5–11.
Messick, S. (1995). Validity of psychological assessment: Validation of inferences of persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749.
Murphy Odo, D. (2012). The impact of high school exit exams on ESL learners in British Columbia. *English Language Teaching, 5*, 1–8.
Norton, B. (2013). *Identity and language learning: Extending the conversation* (2nd ed.). Bristol, UK: Multilingual Matters.
Paris, S, Lawton, T, Turner, J, & Roth, J. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher, 20*, 12–20.
Pollard, A, Triggs, P, Broadfoot, P, McNess, P, & Osborn, M. (2000). *What pupils say: Changing policy and practice in primary education*. London: Continuum.
Qi, L. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education, 14*, 51–74.

Ravitz, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. New York: Basic Book.
Rea Dickens, P, & Scott, C. (2007). Washback from language tests on teaching, learning and policy: Evidence from diverse settings. *Assessment in Education, 14*, 1–7.
Richardson, J. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review, 6*, 135–147.
Roderick, M, & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational and Policy Analysis, 23*, 197–227.
Roessingh, H, Kover, P, & Watt, D. (2005). Developing cognitive academic language proficiency: The journey. *TESL Canada Journal, 23*, 1–27.
Ross, S. (2008). Language testing in Asia: Evolution, innovation and policy challenges. *Language Testing, 25*, 5–13.
Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow: Pearson Education Limited.
Spolsky, B. (2007). On second thoughts. In J Fox, M Wesche, D Bayliss, L Cheng, C Turner, & C Doe (Eds.), *Language testing reconsidered* (pp. 9–18). Ottawa: University of Ottawa Press.
Stobart, G, & Eggen, T. (2012). High-stakes testing – value, fairness and consequences. *Assessment in Education, 19*, 1–6.
Swain, M, & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*, 371–391.
Taguchi, T, Magid, M, & Papi, M. (2009). The L2 Motivational Self System among Japanese, Chinese and Iranian learners of English: A comparative study. In Z Dörnyei & E Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (pp. 66–97). Bristol: Multilingual Matters.
Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Immediate publication on acceptance
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com