

An Evaluation of Alignment between French Language National Test and the Common European Framework of Reference for Languages Using Item Mapping

Preud Bairaman*, Sungworn Ngudgratoke, and Nalinee Na Nakorn

Educational Measurement and Evaluation, School of Educational Studies, Sukhothai Thammathirat Open University, Thailand

*Corresponding author pbairaman@gmail.com

Received 31 July 2017 | Revised 30 December 2017 | Accepted 10 January 2018 | Published 29 November 2018

Abstract

Standards-based education has become a crucial issue with which evaluating the learning of French language in Thailand was then required to comply. Since the alignment is the core idea in systematic and standards-based reform, this study aimed to evaluate the alignment between Thailand's Professional and Academic Aptitude Test in French language (PAT 7.1) and the Common European Framework of Reference for Languages by using item mapping and to study the factors affecting the misalignment between Thailand's Professional and Academic Aptitude Test in French language (PAT 7.1) and the Common European Framework of Reference for Languages. Variables were composed of contents and cognitive demands. Populations were 100 items of PAT 7.1, 388 grade-12 students (Matthayomsuksa 6) in English-French program (academic year 2017) in 10 schools under the South 2's Development Center of French language (Le Centre pour le Développement du Français, CDF Hatyai) and 3 qualified panelists. A sample of 163 students were selected using the multistage sampling method. Research tools consisted of a PAT 7.1 test, two alignment matrixes and a set of open-ended questions. Data was analyzed using the proportion, Rasch IRT model, chi-square, percentage and content analysis.

The research revealed that there were 16 items of PAT 7.1 aligning with the Common European Framework of Reference for Languages. Most of them were classified into A2 level and evaluated analysis/investigate skill. Moreover, item mapping result was coherent with that of panelists' judgement. Three main factors affecting the misalignment were highlighted: the incoherence between item content and standard content, the unmatched levels of cognitive demands and the language usage, including the grammar, in the items that was irrelevant to standard and to students' language proficiency levels.

Keywords: French language, CEFR, alignment, item mapping

■ Introduction

Standards-based education has been crucially focused on as a national issue in Thailand. Concerning policy was established to support the coherence between contents, cognitive demands and instruction (Office of the Basic Education Commission, 2014). Thai scholars declared that alignment and its evaluation should be put into practice for the standards-based education (Ngudgratoke, 2013; Na Nakorn, 2017). To introduce this trend into the system, Ngudgratoke (2013, p. 2) argued that these 4 educative components should be in agreement: standard, curriculum, instruction and evaluation. French language has been taught in the Thai basic education system as a second language for decades. It has been required to match with

standards. Both the Basic Education Core Curriculum B.E. 2551 (A.D. 2008), and the Common European Framework of Reference for Languages (CEFR) have recently become the keys. CEFR has nowadays been used as another important guideline for French language learning, teaching and assessment, as its *raison d'être*, in the country due to the collaborative effort between the Office of Basic Education Commission (OBEC), Thailand's Ministry of Education and French Embassy in Bangkok. Since 1998, DELF (le diplôme d'études en langue française), a CEFR standardized test, has been used to assess Thai students' French language proficiency. An ideal trend is that every student studying French language in the basic education system should be proficient and

DELTA certified. Nevertheless, a number of students attending and achieving the test did not represent the true number of students studying French language in the system. Bairaman, Ngudgratoke and Na Nakorn (2015) revealed that misaligning items in schools' French language tests with CEFR may be a variable of this problem. However, the previous evaluation of alignment between such tests and CEFR depended solely on panelists' judgement. Bhola, Impara and Buckendahl (2003, as cited in Kaira, 2010, p. 5) stated that this kind of method could lead to misleading inferences about students' achievements and growth because the range of difficulty of items was not considered.

This research was therefore to evaluate the alignment by involving the students' ability into the judgement by using the national test of French language, called Professional and Academic Aptitude Test in French language (PAT 7.1), session October B.E. 2558 (A.D. 2015), in order to provide the information useful for the further test development coherent with the standard promoted nowadays.

■ Objectives

1. To evaluate the alignment between Thailand's Professional and Academic Aptitude Test in French language (PAT 7.1) and the Common European Framework of Reference for Languages by using item mapping.

2. To study the factors affecting the misalignment between Thailand's Professional and Academic Aptitude Test in French language (PAT 7.1) and the Common European Framework of Reference for Languages.

■ Theoretical framework

Alignment has been considered as the core idea in systematic and standards-based education (Smith & O'Day, 1991, as cited in Porter, 2002, p. 5), and traditionally an important key of tests and measurements, as known as content validity (Anderson, 2010, p. 255). However, it was deeply defined as the degree of agreement, coherence and match between educational components such as standard and instruction, instruction and assessment and standard and assessment etc. (Webb, 2002; Webb, Horton, & O'Neal, 2002; Roach, Elliott, & Webb, 2005; Squires, 2012). In order to evaluate the alignment, 3 models were invented and frequently used: Webb model, Surveys of Enacted Curriculum (SEC) model and Achieve model (Case, Jorgensen, & Zucker, 2004, pp. 5-11). But, the constraint is that they all depend solely on panelists' judgement. Bhola et al. (2003 as cited in Kaira, 2010, p. 5) stated that the study of alignment without considering students' ability or the difficulty of items could lead to inappropriate result when inferring student's achievement and growth. In order to limit the errors from evaluation of alignment, item mapping was therefore highlighted in this study. Item mapping is a method of standard setting used for assessing whether the students have the knowledge and skills as determined in standard (Cizek & Bunch, 2007; Wang, 2003, pp. 231-232). Kaira (2010, p. 10) and Wang (2003, p. 233) stated that in this method item difficulty will be placed on the same scale, in terms of histogram, by using a criterion like response probability before being passed to panelists for judgement by using students' minimal competency. The theoretical framework is as follows:

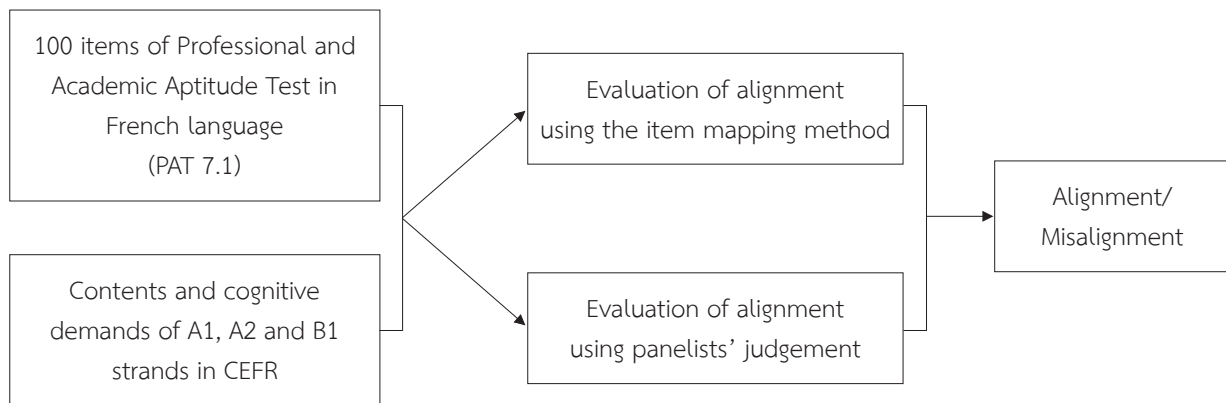


Figure 1 Theoretical framework

As shown in Figure 1, items of PAT 7.1, contents and cognitive demands of CEFR's A1, A2 and B1 were considered as independent variables. Three language levels were chosen for this research as overall learning hours for French language in the secondary school system does not exceed 550 hours. This complies with the hours, estimated by CIEP (Centre international d'études pédagogiques), to be proficient in French language's A1, A2 and B1 (Campus France, 2016). Porter's cognitive demands (recall, demonstrate/explain, analyze/investigate, evaluate and generate/create) were used instead of Bloom's taxonomy as Porter's was specifically proposed for assessing language arts (Hess, 2006, p. 3). In order to find out the alignment together with the misalignment, two methods of item classification (item mapping and panelists' judgement) were conducted and their results were finally compared.

Methodology

Populations in this study were composed of 100 items of PAT 7.1, session October B.E. 2558 (A.D. 2015), 388 grade-12 students (Matthayomsuksa 6) in English-French program (academic year 2017) in 10 schools under the South 2's Development Center of French language (Le Centre pour le Développement du Français, CDF Hatyai) and 3 panelists who were experts in French language teaching and assessment. A sample of 163 students were selected using the multistage sampling method. Research tools constituted a PAT 7.1 test, session October B.E. 2558

(A.D. 2015), two alignment matrixes (the first used for evaluating the alignment between CEFR strands and cognitive demands; the second for evaluating the alignment between the items, competences, contents and cognitive demands), and a set of open-ended questions. Data collection was divided into 4 phases: alignment evaluation by a panel of 3 experts in French language teaching and assessment, 3 days of testing the 163 sample students with 100 items of PAT 7.1, 3 rounds of item mapping with the difficulty of items obtained using the Rasch IRT model, and 1 day of focus group interview with 3 panelists to determine the factors affecting the misalignment.

In the analyzing process the number of items aligning with CEFR, based on 3 panelists' judgement in the first phase, was statistically analyzed using the proportion. Following this, the answers of samples were put into the Rasch IRT model as follows.

$$p(\theta_i) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}}$$

where θ - students' ability
 (from -3 to +3, and $M=0$)
 i - number of item
 a - constant discrimination
 b - item difficulty
 D - constant value equal to 1.7
 e - exponential function

In the consideration of finding, all items were ordered by b_i from the easiest to the hardest (-1.03-1.59). The response probability (RP) of 0.67 was

then set as this criterion has been broadly accepted for educational testing and able to maximize the information carried in the correct response, compared to RP criterion of 0.50 and 0.80 (Kaira, 2010, pp. 39-40; Cizek & Bunch, 2007, p. 162; Huynh, 2006, as cited in Ngudgratoke, Pinyopanuwat, & Na Nakorn, 2007, p. 26). Each panelist examined the item one by one which aligned with CEFR. A1, A2 and B1 cut scores ($A1=0.058$, $A2=0.738$, $B1=1.825$) were subsequently calculated by using the means of θ in each level ($\theta=b_i+0.708$). Items of which θ was inferior to their level's cut score were reconsidered by the panelists whether it was fit to other level or to be eliminated. In order to compare the items aligning with CEFR from both item mapping and panelists' judgment, chi-square was used. The content analysis was conducted for checking the factors affecting the misalignment. Finally, the number of aligning items, based on contents and cognitive demands, was analyzed using the percentage.

Result

From the study, 16 items aligned with CEFR as shown in table 1 and table 2

Table 1 Numbers of items aligning with CEFR

Items	b	Judgements	
		Panelist	Item mapping
43	0.56	A2	A2
45	0.20	B1	A2
51	0.48	A2	A2
55	0.38	A2	A2
56	1.59	A2	B1
57	0.05	A2	A2
59	0.05	A2	A2
60	1.21	A2	B1
63	0.21	A2	A2
64	0.38	A2	A2
65	-0.56	A2	A1

Items	b	Judgements	
		Panelist	Item mapping
66	0.38	B1	A2
70	0.38	A2	A2
72	-0.44	A2	A1
75	0.05	B1	A2
80	0.10	A2	A2

All items aligning with CEFR assessed students' reading competence. A1 items matched content in strand 1 (12.50%), A2 items matched equally strand 1 (25%), strand 2 (25%) and strand 3 (25%), and B1 items matched strand 1 (12.50%). Most items aligning with CEFR assessed analysis/investigate (A1=12.50%, A2=56.25%, B1=12.50%). Another assessed evaluate (12.50%) and demonstrate/explain (12.50%) all of which were classified into A2 level.

Table 2 Items aligning with CEFR

	Levels					
	A1	%	A2	%	B1	%
Item mapping	2	12.50	12	75.00	2	12.50
Panelists	0	0.00	13	81.25	3	18.75

Alignment results from item mapping and panelists' judgement were not different ($X^2=1.231$, $p=0.54$) as shown in table 3

Table 3 Alignment results from items mapping and panelists' judgement

Panelists	Item mapping			X^2	p
	A1	A2	B1		
A1	0	0	0	1.231	0.54
A2	2	9	2		
B1	0	3	0		

Factors affecting the misalignment

The study revealed that 84 items misaligned with CEFR whilst the 7 aligning with CEFR did not match each other when being compared between the 2 methods. The issues contributing to this misalignment were the incoherence between item content and CEFR content, the inappropriate assessment of students' competences when compared to CEFR strands, the unmatched levels of cognitive demands between items and CEFR, the language usage in the items that was irrelevant to CEFR and to students' language proficiency levels (particularly the difficult vocabulary) excess verbiage and complex grammatical structure.

■ Discussion

Evaluation of alignment using item mapping

Due to RP of 0.67, middle criterion ranging from 0.50 to 0.80, the adjacent agreement was highest at A2 level. This was the upper-middle level of grade-12 students (Matthayomsuksa 6). This finding relevant to that of Kaira (2010) indicating when determining RP criterion of 0.67, items in both mathematics and reading matched the higher Educational Functioning Levels (EFL) of Massachusetts Adult Proficiency Test (MAPT), from Low Adult Secondary. That the alignment result, based on chi-square analysis, which did not show the significant difference, was in contrast with the finding of Kaira (2010); this demonstrated that there were general differences between these two methods of item classification in both mathematics and reading, regardless of RP criterion (0.50 or 0.67). Anyway, the close result of this study may be explained by the panelists' French language proficiency, experiences and understanding in item mapping. Kaira (2010, pp. 100-106) argued that an important variable that would affect the similarity or the difference between the result of item mapping and panelists' judgement is the panelists' knowledge and ability of judgement, despite the use of different RP criterions.

Misalignment between items and CEFR

Porter (2004), Näsström (2008), and Polikoff, Porter and Smithson (2011) stated that the incoherence between item content and standard content is a core

variable affecting degree of alignment. When items are neither irrelevant to standard nor do not represent its content, the low degree of alignment will occur. Based on the finding, there were only 16% of items matching CEFR. This revealed that the CEFR contents were not sufficiently involved in PAT 7.1. Näsström (2008, p.26) called this tendency that the mix between standard and freestanding knowledge; it led finally to the misalignment. This issue of content incoherence contributed to the assessments of students' competences inapt with the standard as well. In CEFR, each competence requires specific assessment (listening, speaking, reading, and writing), so the multiple choices in PAT 7.1 were not altogether fit and finally led to the remarkable misalignment. The study of Chairuang (2014) affirmed this reason as he found that the inappropriate assessment practices impacted the misalignment between intended assessment standards and teachers' enacted assessments on Mathematics for grade 10 students (Matthayomsuksa 4) in Chiangmai province, Thailand.

Moreover, the single emphasis on the Basic Education Core Curriculum B.E. 2551 (A.D. 2008) may be another variable explicating the incoherence between item content and CEFR content. Should the content scopes in the Basic Education Core Curriculum B.E. 2551 (A.D. 2008) be studied, the broad description of content was unsurprisingly found. It once covered all 3 levels (Ministry of Education, 2008), or from grade 10-12 (Matthayomsuksa 4-6), while CEFR content was determined separately into different language levels: A1, A2, B1, B2, C1 and C2 (Council of Europe, 2001). Porter described that the standard of which content is detailed in terms of students' levels lured higher alignment than the one covering the entire range of students' level i.e. grade 3-5, grade 6-8 and grade 9-12 etc. (Olson, 2003). Herman and Desimone (2000, as cited in Porter, 2002) mentioned that the different degrees of alignment can be produced by the broad and fine scopes of standard content, similar to Porter's statement,

Another factor affecting the misalignment was the discrepancy of cognitive demand. This related

to empirical evidence presented by Polikoff et al. (2011). They studied 19 states' assessments on English, reading and sciences and indicated that the balance and imbalance of cognitive demands between items and standard may describe the alignment and misalignment. Kaira (2010, p. 107) also discussed that this mismatch of cognitive demands may not only impact the students' frustration and stress when doing the test, but also might affect their subsequent performance in an item.

Finally, that the language scopes of A1, A2 and B1 were not seriously taken into account may have led to the misalignment observed. Analyzed by Eaquels (2015), the scope of each level covers its specific vocabulary and grammar that learners are required to know. Hence, the language usage in the items, regardless of the vocabulary, verbiage and grammatical structure introduced in CEFR, may be an important factor affecting the misalignment found in this research. Kaira (2010, p. 107) pointed out that the irrelevant characteristics of items, i.e. language, could be a variable in explaining the alignment or the misalignment between the test and the standard.

■ Recommendation

Recommendation for using the finding

1. A1, A2 and B1 contents, cognitive demands, and assessments fit to the competences should be seriously studied and included in the subsequent item construction.

2. Evaluating the alignment between items of PAT 7.1 and CEFR should be conducted by trialing with a group of samples before being administered to examinees.

Recommendation for further study

1. Using item mapping to evaluate the alignment between PAT 7.1 and CEFR with students' data from every CDF. Different RP criteria should also be used.

2. Evaluating the alignment between PAT 7.1 and the Basic Education Core Curriculum B.E. 2551 (A.D. 2008).

■ References

- Anderson, L. W. (2010). Curricular alignment: A re-examination. *Theory Into Practice*, 41(4), 255-260. doi:10.1207/s15430421tip4104_9
- Bairaman, P., Ngudgratoke, S., & Na Nakorn, N. (2015). *Effect of alignment between assessment and the common European Framework of reference for languages on French language achievement of Grade-11 students*. Nonthaburi: School of Educational Studies, Sukhothai Thammathirat Open University.
- Campus France. (2016). Les tests et diplômes de français langue étrangère [Brochure]. Retrieved from http://ressources.campusfrance.org/catalogues_recherche/diplomes/fr/fle_fr.pdf
- Case, B. J., Jorgensen, M. A., & Zucker, S. (2004). Alignment in educational assessment. Retrieved from https://images.pearsonassessments.com/images/tmrs/tmrs_rg/AlignEdAss.pdf?WT.mc_id=TMRS_Alignment_in_Educational_Assessment
- Chairuang, N. (2014). *Impact of misalignment between intended assessment standards and teachers' assessment practices on Mathematics Achievement of Mathayom Suksa IV Students in Chiang Mai Province*. (Master of Education in Educational Evaluation Thesis). Sukhothai Thammathirat Open University, Nonthaburi.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: a guide to establishing and evaluating performance standards on tests*. Thousand Oaks, California: SAGE.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Language Policy Unit.
- Eaquels. (2015). *Inventaire linguistique des contenus clés des niveaux du CECRL*. Brian North: Fondation Eurocentres/Eaquels.
- Hess, K. K. (2006). Cognitive complexity: Applying webb DOK levels to Bloom's Taxonomy. Retrieved from <http://www.allen.kyschools.us/Downloads/Cognitive%20Complexity%20Webb%20DOK.pdf>
- Kaira, L. T. (2010). *Using Item Mapping to Evaluate Alignment between Curriculum and Assessment* (Doctor of Education Dissertations). University of Massachusetts-Amherst, Amherst.
- Ministry of Education. (2008). *Basic education core curriculum B.E. 2551 (A.D. 2008)*. Bangkok: Ministry of Education.

- Na Nakorn, N. (2017). *Usability of fair and compatible ONET across years: Social studies, religion and culture subject case*. Nonthaburi: School of Educational Studies, Sukhothai Thammathirat Open University.
- Näsström, G. (2008). *Measurement of alignment between standards and assessment* (Doctor Thesis). Umeå University, Umeå, Sweden.
- Ngudgratoke, S. (2013). Designing balanced assessment systems in a standards-based education system. *STOU Education Journal*, 6(2), 1-8.
- Ngudgratoke, S., Pinyopanuwat, R., & Na Nakorn, N. (2007). Using rasch model and bookmark standard setting procedure to establish a cutscore on STOU-TBS test. *Journal of Research Methodology*, 20(1), 23-41.
- Office of the Basic Education Commission. (2014). *Policies of the basic education commission B.E. 2558*. Bangkok: The Agricultural Co-operative Federation of Thailand.
- Olson, L. (2003). Standards and tests: Keeping them aligned. *Research Points: Essential Information for Education Policy*, 1(1), 1-4.
- Polikoff, M. S., Porter, A. C., & Smithson, J. (2011). How well aligned are state assessments of student achievement with state content standards? *American Educational Research Journal*, 48(4), 965-995. doi:10.3120/0002831211410684
- Porter, A. C. (2002). Measuring the content of instructions: Uses in research and practice. *Educational Research*, 31(7), 3-14.
- . (2004). Curriculum assessment. Retrieved from [http://www.andyporter.org/sites/andyporter.org/files/papers/Curriculum Assessment.pdf](http://www.andyporter.org/sites/andyporter.org/files/papers/Curriculum%20Assessment.pdf)
- Roach, A. T., Elliott, S. N., & Webb, N. L. (2005). Alignment of an alternate assessment with state academic standards. *Journal of Special Education*, 38(4), 218-231.
- Squires, D. (2012). Curriculum alignment research suggests that alignment can improve student achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 85(4), 129-135. doi:10.1080/00098655.2012.657723
- Wang, N. (2003). Use of the rasch IRT model in standard setting: An item-mapping method. *Journal of Educational Measurement*, 40(3), 231-252.
- Webb, N. L. (2002). *An analysis of the alignment between mathematics standards and assessments for three states*. Madison: University of Wisconsin-Madison.
- Webb, N. L., Horton, M., & O'Neal, S. (2002). An analysis of the alignment between language arts standards and assessments for four states. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.527.2693&rep=rep1&type=pdf>