The Canadian Journal for the Scholarship of Teaching and Learning

Volume 3 | Issue 1 Article 5

April 2012

Evaluation of an Internship Assessment Grid for Francophone Physical and Health Education Student Interns

Jaouad Alem
Laurentian University, jalem@laurentian.ca
Céline Boudreau-Larivière
Laurentian University, cboudreaulariviere@laurentian.ca

http://dx.doi.org/10.5206/cjsotl-rcacea.2012.1.5

Evaluation of an Internship Assessment Grid for Francophone Physical and Health Education Student Interns

Abstract

The objective of the present study is to analyze four metric qualities of an assessment grid for internship placements used by professionals to evaluate a sample of 110 Franco-Ontarian student interns registered between 2006 and 2009 at Laurentian University in the School of Human Kinetics. The evaluation grid was composed of 26 criteria. The four metric qualities that were analyzed were: the degree of difficulty, the degree of discrimination, the internal consistency, and the concurrent validity. Each intern's performance was assessed by three individuals: the professional supervisor, the intern (self-assessment) and the university professor who coordinates the internship placement. The analysis of the three assessments based on the Education Testing Service Method indicates that the assessment of the professional supervisors and intern self-assessment are too high (difficulty index, $p_i = 20$) and produced a discrimination power of zero between the interns (discrimination index, $D_i = 0$). The analysis of the internal consistency of the criteria indicates that a number are too highly interrelated (Cronbach's alpha = 0.97) and that ten criteria can be removed from the evaluation grid, as they are redundant. Concurrent validity, determined by calculating three correlations between the three dimensions of the evaluation grid (before, during, and after the teaching session) and the overall rating of the intern, was demonstrated insofar as the lowest correlation between the assessment of the intern's performance and the measurement criterion (overall rating of the intern's performance) was significant ($r\,(106) = .76, p < .001$).

L'objectif de la présente étude est d'analyser quatre qualités métriques de l'évaluation d'un échantillon de 110 étudiants stagiaires franco-ontariens inscrits entre 2006 et 2009 à l'Université Laurentienne à l'École des sciences de l'activité physique. La grille d'évaluation était composée de 26 critères. Les quatre qualités métriques sont: le degré de difficulté, le degré de discrimination, la consistance interne et la validité concomitante des critères. Les stages ont été évalués par trois personnes: le superviseur de stage, le stagiaire (auto-évaluation) et le professeur d'université qui coordonne le stage. L'analyse des items selon la méthode ETS (Educational testing service) indique que les évaluations des superviseurs de stage et l'auto-évaluation des stagiaires sont trop élevées (indice de difficulté, $p_i = 20$) et ne sont pas discriminantes (indice de discrimination $D_i = 0$). L'analyse de la consistance interne des critères indique qu'un certain nombre sont trop fortement corrélées entre eux (coefficient alpha de Cronbach = 0,97) et que dix critères peuvent être retirés de la grille d'évaluation car ils sont redondants. La validité concomitante, déterminée par le calcul de trois corrélations entre les trois dimensions de la grille d'évaluation (avant, pendant, et après la session d'enseignement) et la note globale du stagiaire, a été démontrée dans la mesure où la plus faible corrélation entre l'évaluation de la performance du stagiaire et le critère de mesure (note globale de la performance du stagiaire) était significative ($r\,(106) = .76, p < .001$).

Keywords

validity, analysis of criteria, teaching competency, intern

Cover Page Footnote

This research was supported by Health Canada through a grant from the Consortium national de formation en santé. Correspondence concerning this article should be addressed to Jaouad Alem, School of Human

Kinetics, Laurentian University, 935 Ramsey Lake Road, Sudbury, Ontario, Canada, P3E 2C6. Email: jalem@laurentian.ca

During internship placements in teaching or coaching, francophone students within the School of Human Kinetics at Laurentian University are supervised exclusively by professionals within schools or sports clubs. The intern's performance is assessed by three individuals: the professional supervisor (worth 60 points of their grade), the intern (self-assessment worth 10 points), and the university professor who coordinates the internship placement (final written report on the field placement worth 30 points). In the present study, we have developed an assessment grid to assist professional supervisors to evaluate the student interns. We have also tested the validity of this grid in order to determine how accurate these assessments are in evaluating the performance of the interns.

According to Lessard and Tardif (2005), field placement is a culminating experience in teacher training. The influence of professional supervisors on the careers of interns has been proven: supervisors are perceived as the ones who guide the interns into becoming teachers (McIntyre & Byrd, 1998). Futhermore, interns tend to identify with the supervisors throughout their field placements (Legault, Charbonneau, Chevrier, & Grégoire-Dugas, 1997). However, the correlation between academic success and professional success does not exceed .16 as reported in the meta-analysis of Fraser, Walberg, Welch, and Hattie (1987) and as previously documented for Moroccan and Senegalese intern cohorts in the field of physical education (Alem, Dadouchi, & Kpazai, 2010).

The question of empirically evaluating the performance of teachers was first comprehensively addressed by Medley, Coker, and Soar (1984). According to Spallanzani, Sarrasin, & Goyette (1995) professional supervisors in physical education state that they have had problems assessing the quality of teaching of future teachers in a discriminating manner. This finding was also recently confirmed (Alem & Boudreau-Larivière, 2009). Accordingly, the availability of validated assessment tools may assist professional supervisors to better evaluate interns.

Desbiens (2009) reports that the lack of a verification process of teachers' qualifications, competencies and beliefs on how to become teachers can limit the educational scope of a field placement. Recently, Desbiens (2009) presented some key problematic elements regarding the skills of professional supervisors overseeing teaching interns in Québec. For instance, professional supervisors are rarely chosen based on their supervisory abilities because these abilities are difficult to define and measure. In some environments, the fact that one has been trained has little bearing on being recruited to supervise and is often based on the availability to supervise rather than competency to supervise. A study released in Québec by Lacroix-Roy, Lessard and Garant (2003) indicated that 81.5% of professional supervisors in the province received less than 30 hours of training in supervision, with the average being 14.8 hours. Furthermore, Koster, Korthagen, and Wubbels (1998) report that supervisors do not always have the appropriate tools to teach the concepts, theories and principles related to the field of teaching. This creates a lack of consistency between what is taught in the training program and during the field placement (Mitchell & Schwager, 1993). Finally, professional supervisors are less likely to observe interns systematically and in a sustained manner (Spallanzani et al., 1995). These studies indicate that there is variability in the competencies of the professional supervisors to evaluate their interns.

In the late 80s Westerman (1989) proposed that intern assessment tools should include criteria reflecting the results of recent research on teaching effectiveness, particularly criteria related to the affective domain. According to Bourque (1991), the educational intervention competencies most often used when assessing teaching interns are abilities such as planning, preparation, and utilization of teaching strategies, oral and written communication, knowledge of the teaching content and resources and, finally, class management. Bujold (1997, 2002) maintains that teaching interns will perform better if they are motivated to build upon what

they learned during the placement and to continue to grow professionally through knowledge transfer or even the ability to adapt their learning and apply it in new teaching situations.

The research conclusions outlined above are worth noting as they validate the hypotheses put forth by Shechtman (1989), Byrnes, Kiger, and Shechtman (2000) and Alem (2003) that certain personal abilities have a superior predictive value compared to cognitive abilities for explaining the success of interns during field placements. Shechtman and Godfried (1993) established the following three essential teaching competencies by conducting a factorial analysis of a set of 13 competencies deemed essential: (a) verbal communication, (b) interpersonal relations, and (c) sense of leadership. Other researchers had previously noted the importance of these competencies (see Dunkin & Barnes, 1986; Erdle, Murray, & Rushton, 1985; Guyton & Farokhi, 1987; Lowman, 1984). Furthermore, these competencies contribute to a more accurate prediction of initial success in teaching rather than the candidate's university academic records (Shechtman, 1989). Byrne, Kiger, and Shechtman (2000) have established that these competencies are a good measure of distinct, independent components related to success in teaching.

Another measure of competencies is the overall rating obtained following consensus between the assessors. This overall rating gives a general impression of the teacher candidate's performance during his or her placement. It is a holistic rating measured on a Lickert scale that better predicts initial success in teaching than the grades. In fact, other methods such as personality questionnaires, projective tests, and one-on-one interviews are all less reliable in terms of predicting initial success in teaching (Luther & Lewin, 1991; Shechtman, 1989, 1998; Shields & Daniele, 1982) compared to the holistic rating (i.e., overall rating).

Given these research results as well as the fact that neither the university professors responsible for internship placements within the School of Human Kinetics at Laurentian University, nor the professional supervisors, have an empirically validated assessment grid for their interns, we are proposing the development and validation of a set of criteria in an assessment grid aimed at evaluating the performance of Franco-Ontarian student interns. To address this issue, an assessment grid consisting of 26 criteria was developed by the authors. The authors selected the following four metric qualities to analyze the criteria: (a) the degree of difficulty, (b) the degree of discrimination among interns, (c) the degree of internal consistency, and (d) the concurrent validity.

Method

Developing the Assessment Grid

The assessment grid was developed by five student volunteer interns and four university professors responsible for the field placements. An initial meeting with the student interns was held in the form of a brainstorming session to identify the most relevant criteria in assessing interns. To identify the criteria, participants drew from Dunkin and Biddle's (1974) conceptual model of teaching analysis, which consists of four variables that describe the teaching intern learning process (i.e., presage, context, process, product). In addition, a fifth variable, referred to as the program variable as detailed in Brunelle, Drouin, Godbout and Tousignant (1988), was integrated into the initial model. These five variables have been suggested to highlight the attitudes and skills of trainees (Dunkin & Biddle, 1974; Brunelle, Drouin, Godbout & Tousignant, 1988).

Two professional supervisors (secondary schools) were asked to provide his or her feedback on the final assessment grid. A second meeting of the interns and professors served to short-list the most relevant criteria by drawing from the three conditions proposed by

DOI: http://dx.doi.org/10.5206/cjsotl-rcacea.2012.1.5

Godbout (1988): (a) the importance of the criteria, (b) their interdependence, and (c) their observability. After verifying that the criteria were in line with previous research (e.g., Bourque, 1991; Dunkin & Barnes, 1986; Erdle et al., 1985; Guyton & Farokhi, 1987; Lowman, 1984; Shechtman & Godfried, 1993) the criteria were then selected by the interns and professors based on consensus or even unanimity. A total of 26 criteria were identified as being important to intern assessment (see Table 1 for the list of criteria).

Table 1
The 26 Criteria in the Initial Assessment Grid

Criteria
A. Overall rating
1. Were the voice, volume and tone clear and inspiring?
2. Did the intern create and maintain pleasant and productive sessions?
3. Was the intern able to modify the training/teaching model according to the group's skill level?
4. Was the intern's overall contribution positive?
5. Did the intern project a positive attitude?
6. Was the intern punctual?
7. What was the quality of his or her attire?
8. What was his or her level of respect shown towards the learners?
9. What was his or her level of charisma?
10. Did the intern show an ability to listen, understand the needs expressed and rephrase the questions asked?
B. BEFORE the teaching session
11. How prepared was the intern for each training/teaching session?
12. Were the targeted skills clearly formulated? Were they consistent with the levels of the learners?
13. Were the anticipated exercises progressive (easiest to hardest)? Did they present a good level of challenge?
14. How was the quality of choice and optimal use of the material?
15. Were the session plans clear? Were they illustrated with diagrams?
C. DURING the session
16. Were the teaching/practice sessions well organized, clear and progressive?
17.Control of the classroom group (roll call, explanation of the session, managing materials).
18. Flow of the session itself: clear demonstrations, proper timing, according to needs, demonstrations illustrated using examples.
19. The intern provided the athletes/students with appropriate feedback and made the necessary corrections.
20. The intern was able to get the learners to cooperate.
21. The intern was able to provide optimal time for learning (active learning task).
22. The intern visually scanned the learners regularly and completely (positioned himself/herself in order to observe all learners).
23. At the end of the session: summary of the lesson with the learners.
D. AFTER the session
24. Was the intern able to reflect on his or her actions and be proactive?
25. What was the quality of his or her relationship with the supervisor?
26. What was the quality of the questions he or she asked the supervisor?

The three competencies are verbal communication, interpersonal relations, and sense of leadership. Table 2 provides examples of how the criteria help to measure these competencies.

Table 2
Examples of Criteria within the Evaluation Grid that Assess Verbal Communication, Interpersonal Relations and Sense of Leadership

Verbal communication	Interpersonal relations	Sense of leadership
1. Were the voice, volume and tone clear and inspiring?	3. Was the intern able to modify the training/teaching model according to the group's skill level?	2. Did the intern create and maintain pleasant and productive sessions?
12. Were the targeted skills clearly formulated? Were they consistent with the levels of the learners?	6. Was the intern punctual?	4. Was the intern's overall contribution positive?
19. The intern provided the athletes/students with appropriate feedback and made the necessary corrections.	8. What was his or her level of respect shown towards the learners?	5. Did the intern project a positive attitude?
26. What was the quality of the questions he or she asked the supervisor?	13. Were the anticipated exercises progressive (easiest to hardest)? Did they present a good level of challenge?	7. What was the quality of his or her attire?
	18. Flow of the session itself: clear demonstrations, proper timing, according to needs, demonstrations illustrated using examples.	9. What was his or her level of charisma?
	25. What was the quality of his or her relationship with the supervisor?	17. Control of the classroom group (roll call, explanation of the session, managing materials).
		20. The intern was able to get the learners to cooperate.

DOI: http://dx.doi.org/10.5206/cjsotl-rcacea.2012.1.5

The Variables Studied

The assessment grid consists of four variables: (a) the overall rating of the intern's performance which is assessed using the ten criteria outlined in Part A of Table 1, (b): the intern's level of preparation prior to the intervention session which is assessed using the five criteria outlined in Part B of Table 1, (c) the quality of interaction with his or her environment during the session which is assessed using the eight criteria outlined in Part C of Table 1, and (d) the quality of the intervention immediately after the session which is assessed using the three criteria outlined in Part D of Table 1. The criteria were assessed using Likert's six-level scale to avoid the selection of middle-point ratings by respondents (1=*Poor*, 6= *Outstanding*).

Sample of Student Interns

This research was accredited by the Research Ethics Board from Laurentian University. The database did not record the name of the institution or organization where the student completed his or her internship, the name of the placement supervisor or the names of trainees but it did record the participants' gender.

The sample consisted of 110 Francophone interns (51 females and 59 males) registered in the francophone Physical and Health Education program in the School of Human Kinetics at Laurentian University between 2005 and 2009. During this period, a total of 26 field placements (4^{th} year, 120 hours internship) or practicums (3^{rd} year, 80 hours internship) were undertaken in a teaching context and 84 in a coaching context. For both internship types (field placement and practicum), the assessment grid was provided at the beginning of the internship and was used by the professional supervisor to evaluate each student once at the end of his or her placement. The supervisors were also invited to rate the assessment grid and provide feedback on its usefulness at the end of placement.

Results

To establish the reliability and validity of the Assessment Grid, four metric qualities were analyzed: (a) the internal consistency of the criteria is estimated by Cronbach's alpha coefficient (Cronbach, 1951), (b) support for the concurrent validity was established by correlating the three criterion subscales (i.e., before, during and after the intern's intervention) with the overall rating of the intern's performance during the placement (Shechtman, 1989, 1998; Shechtman & Godfried, 1993), and finally the (c) Difficulty index (p_i) and (d) Discrimination inde*x* (D_i) were evaluated by assessing the criteria based on the Educational Testing Service (ETS) method. This system is based on the sum of 1) the number among the ten participants who obtained the highest marks on the test (10++) and met the criterion and 2) the number among the ten participants who obtained the lowest marks (10--) but met the criterion. The p_i index is the sum of these two statistics whereas the D_i index is the difference between these two statistics. Success in meeting a criterion is defined in the present study as 60%. Therefore a criterion is considered as being met if the participant obtains the following minimum marks: 18/30 (from the university professor); 36/60 (from the professional supervisor); 6/10 (student self-assessment). According to the ETS method, an index of difficulty greater than 17 indicates that the criterion is too simple whereas an index lower than ten indicates that the criterion is too difficult. A criterion for which the degree of discrimination is lower than three indicates that the criterion cannot be used to differentiate between the subjects (Guay, 2000).

Descriptive Analysis

As seen in Table 3, even without the assessment from the university internship coordinator (M = 21.49, SD = 4.98), the assessments from the professional supervisor and from the student (self-assessment) are rated very highly. This suggests that professional supervisors in particular do not take full advantage of the range of the proposed evaluation scales and consequently do not use the assessment grid in an optimal manner.

Table 3
Descriptive Analysis

	N	*Min*	*Max*	*M*	*SD*
Overall grade from the professional supervisor /60 points	110	42	60	56.06	3.32
Intern self-assessment: justified grade/10 points	107	7	10	9.01	.72
Official grade from the university professor (internship coordinator placement official grade: written report /30 points	105	8.50	30	21.49	4.98
Final grade/100 points	105	66	98	86.65	6.23
Overall rating of the intern /6 points	110	3.90	6	5.55	.53
Intern's degree of preparation before the interventions/6 points	110	2	6	5.31	.78
Intern's performance during the intervention/6 points	108	2.25	6	5.42	.67
Intern's performance immediately following the intervention /6 points	110	3	6	5.54	.72
Valid *N* (listwise)	103				

Internal Consistency of the Criteria

Robinson, Shaver, and Wrightsman (1991) as well as Clark and Watson (1995) propose a minimal acceptable Cronbach's alpha coefficient threshold of .80. As seen in Table 4, the internal consistencies of the criteria are all greater than .90 supporting the reliability of the criteria.

Table 4

Internal Consistency of the Criteria

Questionnaire criteria	Cronbach's alpha
All criteria (26 criteria)	.97
Overall rating (10 criteria)	.91
Before the intervention (5 criteria)	.92
During the intervention (8 criteria)	.94
After the intervention (3 criteria)	.92

DOI: http://dx.doi.org/10.5206/cjsotl-rcacea.2012.1.5

Concurrent Validity of the Criteria

Table 5 presents the concurrent validity being defined as the correlations between the assessment of the intern's performance before, during, and after his or her intervention with the overall rating of the intern's performance. Because the correlations are all large, this supports the concurrent validity of the criteria.

Table 5
Concurrent Validity of the Criteria

	Assessment of the intern before the intervention	Assessment of the intern during the intervention	Assessment of the intern after the intervention
Overall rating of the intern's performance	$r\,(108) = .79, p < .001$	$r\,(106) = .76, p < .001$	$r\,(108) = .81, p < .001$

Degree of Difficulty and of Discrimination of the Criteria

Table 6 presents the degree of difficulty (p_i) and degree of discrimination (D_i) for the assessment grid. Using the ETS method outlined above, it appears that the assessments by the professional supervisors and the student self-assessments are too high and present a weak level of discrimination between internship placements. This is consistent with our previous analysis (Alem & Boudreau-Larivière, 2009).

Table 6
Analysis of the Three Intern Assessments Based on the ETS Method

	10++	10--	p_i	D_i
Assessment of the professional supervisor /60	10	10	20	0
Intern self-assessment /10	10	10	20	0
Assessment of the internship report by the university professor /30	10	7	17	3
Final assessment / 100	10	10	20	0

Data presented in Table 7 suggests that 10 of the 26 criteria are problematic based on the ETS method and were eliminated (the problematic criteria are presented in bold).

Table 7
Analysis of the 26 Criteria Based on the ETS Method

		cr1	cr2	cr3	**cr4**	**cr5**	**cr6**	**cr7**	**cr8**	cr9	**cr10**
Overall rating	10++	10	10	10	**10**	**10**	**10**	**10**	**10**	10	**10**
	10--	6	6	6	**8**	**10**	**9**	**8**	**9**	7	**8**
	p_i	16	16	16	**18**	**20**	**19**	**18**	**19**	17	**18**
	D_i	4	4	4	**2**	**0**	**1**	**2**	**1**	3	**2**
		cr11	cr12	cr13	cr14	cr15					
Before	10++	**10**	10	10	10	10					
	10--	**8**	5	5	4	5					
	p_i	**18**	15	15	14	15					
	D_i	**2**	5	5	6	5					
		cr16	cr17	cr18	cr19	**cr20**	**cr21**	cr22	cr23		
During	10++	**10**	10	10	10	**10**	**10**	10	10		
	10--	**9**	6	7	7	**9**	**9**	6	6		
	p_i	**19**	16	17	17	**19**	**19**	16	16		
	D_i	**1**	4	3	3	**1**	**1**	4	4		
		cr24	cr25	cr26							
After	10++	10	10	10							
	10--	5	7	5							
	p_i	15	17	15							
	D_i	5	3	5							

The 16 criteria in the final assessment grid are provided in Table 8.

DOI: http://dx.doi.org/10.5206/cjsotl-rcacea.2012.1.5

Table 8
The 16 Criteria in the Final Assessment Grid

Criteria
A. Overall rating
1. Were the voice, volume and tone clear and inspiring?
2. Did the intern create and maintain pleasant and productive sessions?
3. Was the intern able to modify the training/teaching model according to the group's skill level?
9. What was his or her level of charisma?
B. BEFORE the teaching session
12. Were the targeted skills clearly formulated? Were they consistent with the levels of the learners?
13. Were the anticipated exercises progressive (easiest to hardest)? Did they present a good level of challenge?
14. How was the quality of choice and optimal use of the material?
15. Were the session plans clear? Were they illustrated with diagrams?
C. DURING the session
17. Control of the classroom group (roll call, explanation of the session, managing materials).
18. Flow of the session itself: clear demonstrations, proper timing, according to needs, demonstrations illustrated using examples.
19. The intern provided the athletes/students with appropriate feedback and made the necessary corrections.
22. The intern visually scanned the learners regularly and completely (positioned himself/herself in order to observe all learners).
23. At the end of the session: summary of the lesson with the learners.
D. AFTER the session
24. Was the intern able to reflect on his or her actions and be proactive?
25. What was the quality of his or her relationship with the supervisor?
26. What was the quality of the questions he or she asked the supervisor?

The internal consistency of the 16 criteria retained, as well as the correlations between them, remains positive and significant (see Table 9).

Table 9
Cronbach's alpha coefficient (α) of the Criteria Before and After Removing Ten Criteria

α **before removing criteria**		α **after removing criteria**	
All criteria (26 criteria)	.97	All criteria (16 criteria)	.96
Overall rating (10 criteria)	.91	Overall rating (4 criteria)	.86
Before the intervention (5 criteria)	.92	Before the intervention (4 criteria)	.91
During the intervention (8 criteria)	.94	During the intervention (5 criteria)	.91
After the intervention (3 criteria)	.92	After the intervention (3 criteria)	.92

To verify if we could effectively delete these ten criteria from the questionnaire, we compared two models of sequential regression analysis to predict the overall rating of the intern: the first model with the combined 26 criteria and the second model that considers only the sixteen criteria. Table 10 presents the results of the two sequential regression models in terms of the coefficient of determination (R^2) and the retained criteria to predict the overall rating of the intern.

Table 10
Contribution of the Criteria in terms of Explained Variance Needed to Predict the Overall Rating of the Intern's Performance based on the Two Sequential Regression Models (Original 26 criteria versus 16 retained criteria)

Original criteria	**Retained criteria**
1st sequential regression model (26 criteria) R^2= .77, F(3, 88)=99.48, p < .001	2nd sequential regression (16 criteria) R^2= .72, F(4,88)= 57.04, p < .001
Were the exercises to be completed progressive in nature? Did they present an acceptable level of challenges? β=.34, p < .001 (criterion 13)	Were the exercises to be completed progressive in nature? Did they present an acceptable level of challenges? β=.29, p < .001 (criterion 13)
What was the quality of the relationship between the intern and the professional supervisor? β=.41, p < .001 (criterion 25)	Were the targeted skills clearly formulated? Were they consistent with the learners' levels? β=.26, p < .01 (criterion 12)
Were the targeted skills clearly formulated? Were they consistent with the learners' levels? β=.27, p < .001 (criterion 12)	What was the quality of the questions he or she asked the supervisor? β=.21, p < .01 (criterion 26)
	Flow of the actual session: clear demonstrations, properly timed, according to needs, demonstrations illustrated using examples β=.22, p < .05 (criterion 18)

The two regression models are statistically significant. Even though the first regression model provides a higher percentage of explained variance (77%) than the second (72%), the second regression model is better supported by the conceptual model of Dunkin and Biddle (1974) in that it includes one criterion, criterion 18, which assesses the intern's performance *during* the intervention with the classroom group.

Discussion

In the present investigation, an assessment grid consisting initially of 26 criteria was developed and empirically validated. Criteria to measure the performance of student interns before, during, and after their interventions as well as criteria to measure their overall performance were included in the grid. The ultimate objective of the creation of this grid was to provide a validated assessment grid to the university professors responsible for internship placements as well as the internship supervisors to better score the performance of student interns. The authors selected the following four metric qualities to analyze those criteria: their degree of difficulty, their degree of discrimination, their degree of internal consistency and their concurrent validity. The analysis of the internal consistency of the 26 original criteria revealed that some criteria were too redundant as shown in Table 9 (α = .97). Concurrent validity was demonstrated insofar as the lowest correlation between the measured criterion (i.e., overall rating of the intern) and the 16 criteria in the assessment grid were significant. Analyses of the three assessor groups (i.e., professional supervisor, intern self-assessment, and university professor), based on the Education Testing Service method, indicated that the assessments of the professional supervisors were too high (p_i = 20) and

DOI: http://dx.doi.org/10.5206/cjsotl-rcacea.2012.1.5

produced a discrimination power of zero between the student interns (D_i = 0). These results were not unexpected and in fact support previous findings reported by others such as Desbiens (2009). Furthermore, these results (p_i=20 and D_i=0) apply equally to the self-assessment scores attributed by the student interns.

Our findings highlight that professional supervisors in addition to student interns, at least in our sample, attributed inflated scores that made it difficult to discriminate between strong, average and poor performing interns. Luhanga, Yonge, and Myrick (2008) similarly reported that placement supervisors of nurses in training had difficulty attributing low scores to interns that would result in interns failing their placement. Collectively, it is therefore important to sensitize professional supervisors and student interns to take advantage of the full range of the evaluation scale to better discriminate between the interns' performance.

The analysis of the 26 criteria based on the ETS method indicated that ten criteria could be eliminated from the assessment grid. By comparing the two regression models (the first with 26 criteria, the second with 16 criteria) to predict the overall rating of the intern, we obtained prediction models that were quite similar. We, therefore, believe that the shorter assessment grid consisting of 16 criteria (see Table 8) represents a useful validated instrument to evaluate the performance of student interns during their placements. The internal consistency of the 16 criteria retained, as well as the correlations between them, remains positive and significant.

Future Research and Practical Implications

We expect to perform additional statistical analysis on the assessment grid for other cohorts of interns during the next few years to verify whether the revised assessment grid with fewer criteria possesses similar or superior metric qualities. Creating a shorter assessment grid will serve as a useful tool for professional supervisors to quickly and accurately evaluate the performance of the student intern. Furthermore, the optimized assessment grid could be used by the interns themselves to better self-assess their performance.

Conclusion

The metric qualities of an evaluation grid composed of 26 criteria created to assess the performance of Franco-Ontarian physical and health education students during their internship placement were analyzed. Ten criteria were removed from the evaluation grid, as they were redundant with other criteria. The assessment scores of the professional supervisors and intern self-assessment were too high which greatly limited the capacity to discriminate the performance of interns. The availability of this empirically validated assessment grid will improve the capacity of the professional supervisors and the student interns to provide more accurate evaluations.

References

Alem, J. (2003). *La valeur de l'appréciation par simulation (APS) pour prédire le succès initial en enseignement des candidats aux études en éducation.* Unpublished doctoral dissertation, Université Laval, Sainte-Foy, Canada.

Alem, J., & Boudreau-Larivière, C. (2009). *Les difficultés relatives à l'évaluation des stages pratiques en enseignement des stagiaires franco-ontariens inscrits dans un programme d'éducation physique et santé.* Paper presented at the Symposium RIFEF Nantes, France.

Alem, J., Dadouchi, F., & Kpazai, G. (2010). La prédiction de l'engagement professionnel des étudiants-stagiaires en éducation physique et sportive selon leur engagement académique, leur moyenne académique annuelle et selon leur compétence essentielles en enseignement: Le cas des stagiaires de l'École Normale Supérieure de Casablanca. In C. Amans-Passaga, N. Gal-Petifaux, P. Terral, M. Cizeron, & M. F. Carnus (Eds.), *L'intervention en sport et ses contextes institutionnels: Culture et singularité de l'action* (pp. 121-126). Location: Édition Presses du Centre Universitaire Champollion.

Bourque, L. (1991). *Les critères de compétence dans l'évaluation des stagiaires en formation des maîtres* Essaie de maîtrise, Université Laval, Québec.

Brunelle, J., Drouin, D., Godbout, P., & Tousignant, M. (1988). *La supervision de l'intervention en activité physique*. Montreal, QC: Gaëtan Morin.

Bujold, N. (1997). *Baccalauréat en enseignement préscolaire et primaire. Rapport annuel (1996-1997) présenté au conseil de la Faculté des sciences de l'Éducation.* Laval University, Québec, QC, Canada.

Bujold, N. (2002). La supervision pédagogique. Vue d'ensemble. In M. Boutet & N. Rousseau (Eds.), *Les enjeux de la supervision pédagogique des stages* (Vol. 4, pp. 9-23). Sainte-Foy, QC: Presses de l'Université du Québec.

Byrnes, D. A., Kiger, G., & Shechtman, Z. (2000). Evaluating the use of group interviews to select students into teacher-education programs. *Journal of Teacher Education,* 54(2), 163-172. http://dx.doi.org/10.1177/0022487102250310

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319. http://dx.doi.org/10.1037/1040-3590.7.3.309

Cortina, J. M. (1993). *What is coefficient alpha? An examination of theory and applications* (Vol. 78). Washington, DC,: American Psychological Association.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Orland, FL: Holt, Rinehart and Winston.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 31*, 93-96.

Desbiens, J. F. (2009). *Bien enseigner ne s'apprend pas tout seul!* Paper presented at the Séminaire pensée et culture, Laurentian University, Sudbury, ON, Canada.

Dunkin, M. J., & Barnes, J. (1986). Research on teaching in higher education. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 754-777). New York: Macmillan.

Dunkin, M. J., & Biddle, B. (1974). *The study of teaching*. New York: Holt, Reinhart and Winston.

Erdle, S., Murray, H. G., & Rushton, P. J. (1985). Personality classroom behavior and student ratings of college teaching effectiveness: A path analysis. *Journal of Educational Psychology, 77*, 394-407. http://dx.doi.org/10.1037/0022-0663.77.4.394

DOI: http://dx.doi.org/10.5206/cjsotl-rcacea.2012.1.5

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research, 11*(2), 147-252. http://dx.doi.org/10.1016/0883-0355(87)90035-8

Godbout, P. (1988). Les prélèvements d'information en activité physique: La mesure. In G. Morin (Ed.), *La supervision de l'intervention en activité physique* (pp. 233-299). Montréal: Gaëtan Morin Éditeur.

Guay, M. (2000). Évaluation physique et en sport. *Course Notes EDPH 2407, Physical Health and Education Program*. Laurentian University, Sudbury, ON, Canada.

Guyton, E., & Farokhi, E. (1987). Relationships among academic performance, basic skills, subject matter knowledge, and teaching skills of teacher education graduates. *Journal of Teacher Education, 38*(5), 37-42. http://dx.doi.org/10.1177/002248718703800508

Koster, B., Wubbels, T., & Korthagen, F. A. J. (1998). Is there anything left for us? Functions of cooperating teachers and teacher educators. *European Journal of Teacher Education, 21*(1), 75-89. http://dx.doi.org/10.1080/0261976980210108

Lacroix-Roy, F., Lessard, M., & Garant, C. (2003). *Étude sur les programmes de formation à l'accompagnement des stagiaires.* Quebec: Table MEQ-universités.

Legault, J. P., Charbonneau, B., Chevrier, J., & Grégoire-Dugas, T. (1997). Les besoins de perfectionnement des enseignants associés : approche définitoire participative. In M. Tardif & H. Ziarko (Eds.), *Continuités et ruptures dans la formation des maîtres au Québec* (pp. 90-108). Sainte-Foy, QC: Les Presses de l'Université Laval.

Lessard, C., & Tardif, M. (2005). *La profession d'enseignant aujourd'hui: Evolution, perspectives et enjeux internationaux*. Brussels, Belgium: De Boeck.

Lowman, J. (1984). *Mastering the techniques of teaching*. San Francisco, CA: Jossey-Bass.

Luhanga, F., Yonge, O., & Myrick, F. (2008). "Failure to fail": Issues with grading the unsafe student. *International Journal of Nurse Education Scholarship, 5*(1), 1-14. http://dx.doi.org/10.2202/1548-923X.1366

Luther, M., & Lewin, L. (1991). *Teacher education admissions procedures at York University.* Paper presented at the Fourth Invitational Conference of the Canadian Association of Teacher Education, Kingston, ON, Canada.

McIntyre, D. J., & Byrd, D. M. (1998). Supervision in teacher education. In G. R. Firth & E. Pajak (Eds.), *Handbook of research on school supervision* (pp. 409-427). New York: Macmillan.

Medley, D. M., Coker, H., & Soar, R. S. (1984). *Measurement-based evaluation of teacher performance: An empirical approach*. New York: Longman.

Mitchell, M. F., & Schwager, S. (1993). Improving the student teaching experience: Looking to the research for guidance. *Physical Educator, 50*(1), 31-38.

Organ, D. W., & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology, 48*(4), 775-802. http://dx.doi.org/10.1111/j.1744-6570.1995.tb01781.x

Robinson, J. P., Shaver, P. R., Wrightsman, L. S., & Andrews, F. M. (1991). *Measures of personality and social psychological attitudes*. San Diego, CA: Academic Press.

Shechtman, Z. (1989). The contribution of interpersonal behavior evaluation to the prediction of initial teaching success: A research note. *Teaching and Teacher Education, 5*(3), 243-248. http://dx.doi.org/10.1016/0742-051X(89)90007-3

Shechtman, Z. (1998). Agreement between lay participants and professional assessors: Support of a group assessment procedure for selection purposes. *Journal of Personnel Evaluation in Education, 12*(1), 5-17. http://dx.doi.org/10.1023/A:1007932515469

Shechtman, Z., & Godfried, L. (1993). Assessing the performance and personal traits of teacher education students by a group assessment procedure: A study of concurrent and construct validity. *Journal of Teacher Education, 44*(2), 130-138. http://dx.doi.org/10.1177/0022487193044002007

Shields, J. J. Jr., & Daniele, R. (1982). *Teacher Selection and Retention: A Review of the Literature*. unpublished manuscript.

Spallanzani, C., Sarrasin, J., & Goyette, R. (1995). La perception des enseignants à l'égard de leur rôle de formateur auprès de stagiaires universitaires. *Paper* , *Actes du colloque de l'AQUFOM* (pp. 133-144). McGill University, Montreal, QC, Canada

Westerman, J. E. (1989). Minimum state teacher certification standards and their relationship to effective teaching: Implications for teacher education. *Action in Teacher Education, 11*(2), 25-32. http://dx.doi.org/10.1080/01626620.1989.10462721

DOI: http://dx.doi.org/10.5206/cjsotl-rcacea.2012.1.5