



Canadian Journal of Higher Education
Revue canadienne d'enseignement supérieur
Volume 37, No. 2, 2007, pages 69 - 88
www.ingentaconnect.com/content/csshe/cjhe

Health Sciences Graduate Students' Perceptions of the Quality of their Supervision: A Measurement Scale

Gina Bravo, Julie Saint-Mleux & Marie-France Dubois
University of Sherbrooke

Advisory Committee on Graduate Supervision¹
University of Sherbrooke

ABSTRACT

We developed and evaluated the G3S-SP, a scale measuring health sciences graduate students' perceptions of the quality of their supervision. The scale was developed from a literature review and existing questionnaires. Feedback from health sciences graduate students and supervisors led to a revised version of the scale that was mailed to 215 students enrolled in eight programs of the Faculty of Medicine and Health Sciences at the University of Sherbrooke. Analyses show that mean satisfaction scores differ significantly across programs ($p=0.036$), which supports the scale's discriminant validity. Factor analysis revealed a two-factor structure accounting for 84% of the variance. The first factor ($\alpha=0.88$) assesses the supervisors' involvement in the design and conduct of the student's research project, while the second ($\alpha=0.76$) refers to student-supervisor relationships. We conclude that the G3S-SP is a valuable tool to monitor health sciences graduate students' perception of the quality of their supervision and to identify areas that need improvement.

¹ The Committee included two graduate student representatives (Nathalie Bier and Estelle Vallée) and three University of Sherbrooke faculty members (Sylvie Bourque, Sonia Morin, and Denise St-Cyr Tribble).

RÉSUMÉ

Nous avons développé puis évalué l'échelle G3S-SP, laquelle mesure la qualité de l'encadrement aux études supérieures telle qu'elle est perçue par les étudiantes et étudiants en sciences de la santé. L'échelle a été construite à partir de la littérature et de questionnaires existants, puis révisée suite aux commentaires d'étudiants et de directeurs de recherche en sciences de la santé. La version finale a été postée à 215 personnes étudiant à l'un des huit programmes de la Faculté de médecine et des sciences de la santé de l'Université de Sherbrooke. L'analyse démontre que les scores moyens de satisfaction diffèrent significativement selon les programmes ($p=0,036$), un résultat qui témoigne de la validité discriminante de l'échelle. L'analyse factorielle révèle une structure à deux facteurs expliquant 84 % de la variance. Le premier facteur ($\alpha=0,88$) évalue l'implication du directeur dans la conception et la réalisation du projet de recherche étudiant, tandis que le second ($\alpha=0,76$) réfère aux relations étudiant-directeur. Nous concluons que l'échelle G3S-SP est utile pour décrire la perception qu'ont les étudiantes et étudiants en sciences de la santé de la qualité de leur encadrement et identifier les améliorations qui s'imposent.

INTRODUCTION

Academic departments that offer graduate programs usually require faculty members to supervise research trainees, in addition to teaching and pursuing their own research activities. Moreover, granting agencies often make the awarding of research grants conditional upon training future researchers.

Supervising graduate students is a complex task for which few faculty members have formal training when hired (Brown & Atkins, 1988; Pole, 1998). The vast majority of new faculty members have neither been trained to teach at the university level. However, having students fill out course evaluations is now a widespread practice in universities, which when coupled with individualized counselling, has been shown to improve teaching skills (March & Roche, 1993, 1997; March, Rowe, & Martin, 2002). Research productivity is also assessed periodically, such as when faculty apply for research funds or submit manuscripts for publication. Unlike teaching and research, however, graduate supervision is not systematically evaluated. Moreover, as March, Rowe, and Martin (2002, p. 318) pointed out, university academics typically receive even less training – and have been exposed to even fewer role models – in how to be effective supervisors than in how to be effective classroom teachers.

Graduate supervision can be conceptualized as a dyadic relationship between students and supervisors, each of whom bear some responsibility for the successful outcome of the relationship. Investigating one of its important components, namely graduate students' perspectives on the quality of their

supervision, requires a survey instrument with sound measurement properties for collecting the data. Although there is a vast body of research literature on undergraduate students' evaluations of classroom teaching effectiveness (e.g., March & Roche, 1993, 1997; McKeachie, 1997) and some research on the quality of supervision provided at the graduate level (e.g., Anderson & Swazey, 1998; Hockey, 1995; Holdaway, 1996; Pearson, 1996), there is little empirical research on the reliability and validity of instruments developed to measure graduate students' perceptions of the quality of their supervision (March, Rowe, & Martin, 2002). We aimed at reducing this gap by 1) developing an instrument to assess the subjective experience of supervision by health sciences graduate students, and 2) evaluating its measurement properties on a sample of those students. The paper begins with a brief overview of the literature on this topic, followed by a description of the research design and main findings of the study. We conclude with a discussion of the strengths and weaknesses of the study as well as the potential uses of student assessments.

The Need for a Survey Instrument

The quality of the relationship that graduate students have with their research supervisors can have a major impact on their success, both in obtaining their degrees and in starting their careers (Ferrer de Valero, 2001; Tinto, 1993). According to recent reports from Canada (Canadian Association for Graduate Studies, 2004), the United States (Denecke, 2005) and United Kingdom (Higher Education Funding Council for England, 2005), completion rates in PhD programs have fallen as low as 70% in most disciplines and even lower in the arts and humanities. Meanwhile, time-to-degree has increased considerably, particularly in the arts, humanities, and social sciences (Canadian Association for Graduate Studies, 1997; Elgar, 2003; Ferrer de Valero, 2001; Henderson, Clark, & Reynolds, 1996). Many factors contribute to the ability of students to complete their degree requirements in a timely manner; these include personality traits, work habits, motivation, commitment, and self-directed learning skills (Ferrer de Valero, 2001; Knowles, 1975; Ramos, 1994; Tluczek, 1995). As noted by Elgar (2003) and Ferrer de Valero (2001), not all students have the self-discipline required to adjust from highly-structured undergraduate education to less structured and more individualized graduate studies.

While acknowledging that students have a major responsibility for their own education, many scholars maintain that the high drop-out rate from graduate programs and the increase in the time students spend earning their degrees are partly attributable to poor supervision (Adam, 2002; Association for Support of Graduate Students, 1993; Brown & Atkins, 1988; Elgar, 2003; Farr, 2002; Hahs, 1998; Hinchey & Kimmel, 2000; Kelly, 1998; Kerlin, 1995; Lapidus, 1997; Lovitts, 2001; National Research Council, 1995; Nerad & Cerny, 1999; Nyquist, 2002; Ramos, 1994; Tluczek, 1995). A supervisor who is rarely available or fails to provide constructive and timely feedback to students will likely slow the progress of the student's research work. A supervisor's lack of interest

in student research projects and lack of respect for intellectual property are other indicators of inadequate supervision.

In light of its importance, many graduate schools have become interested in the relationship between students and their research supervisors, as clearly demonstrated by a quick glance at university newspapers and web sites. This increased awareness has led some institutions to adopt specific policies related to supervision (Donald, Saroyan, & Denison, 1995; Holdaway, 1994) or to develop best-practice guidelines that outline the rights and responsibilities of both research trainees and supervisors (Council of Graduate Schools in the United States, 1990; ESRC, 1994; SERC, 1992). Still others have developed thesis-supervision workshops for supervisors, a practice more common in British universities than in their Canadian and American counterparts (Elgar, 2003).

A scale for periodically assessing the quality of graduate supervision, as perceived by students, would be a valuable tool in gauging the impact of such initiatives. Student evaluations would also provide informative feedback on supervisor strengths and weaknesses that could improve supervision. Additionally, if students were invited to complete the scale while still in active training, it could be used to identify and resolve problem situations, provided, of course, that the students agree to identify themselves. In a best-case scenario, this kind of scale would be used to encourage dialogue between students and supervisors and to structure a frank discussion regarding the sources of satisfaction and dissatisfaction on both sides.

In their general surveys on student profiles and satisfaction, some academic institutions include a few questions about the relationship with research supervisors. Most of these questions, however, are very general in nature, asking students for an overall evaluation of the support they get from their supervisors. Consequently, these initiatives do not focus student attention on discrete components of graduate supervision that may vary in quality. Other questions are purely factual, pertaining, for example, to the frequency of meetings between students and supervisors. In addition, few questionnaires have been submitted to rigorous validation studies and their findings reported in peer-reviewed journals. One exception is the Postgraduate Research Experience Questionnaire (PREQ), a survey instrument developed by the Graduate Careers Council of Australia to measure the extent to which recent graduates were satisfied with their supervision. The systematic development of the PREQ, which involved a diverse group of stakeholders, provides strong support for its content and face validity, and results based on two large data collections support its psychometric characteristics (Australian Council for Education Research, 2000; Marsh, Rowe, & Martin, 2002). However, only a few PREQ items are aimed directly at supervisors; the remainder refer more to the academic unit or entire university. In response to this lack of an appropriate instrument, we developed and explored the validity of the Graduate Studies Supervision Scale – Student Perception (G3S-SP) – a self-administered questionnaire specifically designed to assess health sciences graduate students' perceptions of the quality of supervision that they receive from their supervisors.

METHODS

Developing a Preliminary Version of the Scale

The G3S-SP was developed in French and its content translated into English for the purposes of the present paper. Its items were generated from the scientific literature on graduate supervision and related constructs (Anderson & Shannon, 1988; Burlew, 1991; Green & Bauer, 1995; Jacobi, 1991; Roberts & Sprague, 1995; Rose, 2003; Winston & Polkosnik, 1984), the authors' experiences with supervising graduate students, and existing instruments posted on university web sites. To maximize content validity, items were chosen to cover the entire supervision process, from getting students settled in their research environment to designing their research projects and publishing the results. Most items are rated on a 4-point Likert-type scale, with space available for comments and a box for checking "Does not apply," as illustrated below. Although we provided a "Does not apply" option, we expected the majority of items to apply to most graduate students, irrespective of their stage in the program.

My supervisor's feedback contributes to the advancement of my project.					
	Completely disagree	Somewhat disagree	Somewhat agree	Completely agree	Does not apply
Supervisor A	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supervisor B	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Supervisor C	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Comments:					

All items are stated in such a way that higher ratings reflect a greater degree of satisfaction. Students may have more than one supervisor, with distinct roles and complementary expertise, especially in the natural and health sciences where joint supervision is common (Pole, 1998; Pole, Sprokkereef, & Burgess, 1997). Since a given student's satisfaction with any given item could vary across supervisors, students were asked to rate each of their supervisors separately, when applicable. In this paper, "supervisor" refers to a faculty member formally involved in graduate student supervision, regardless of his or her true degree of involvement and with no distinction made between principal supervisor and co-supervisor. The proper term to use could vary across academic institutions.

The G3S-SP was embedded within a larger questionnaire designed to investigate its validity. Four introductory questions, relating respectively to how prepared the students were for graduate studies, their expectations of their programs and research supervisors, and their knowledge of the resources available in the event of a disagreement with a supervisor, preceded the 19 items of the G3S-SP

dealing specifically with supervision quality. These were followed by a series of questions asking respondents, for example, whether they would recommend their supervisors to other students and whether having more than one supervisor caused any particular problems. Other questions focused on determining if respondents wanted to meet with the graduate program administrator or their student representatives to discuss the quality of their supervision. The last section of the questionnaire collected sociodemographic information about the respondents and suggestions for improving the current version of the scale. Since one purpose of the scale was to identify conflicts, the preliminary version of the questionnaire asked students to identify both themselves and their supervisors.

Seeking feedback from health sciences graduate students and supervisors

The content of the preliminary version of the questionnaire was discussed with 12 students in focus groups and six supervisors reached by phone. Master's and PhD students formed two distinct focus groups because we expected them to have differing expectations of their supervisors. Indeed, master's students often want more support in designing their research protocol and more constructive feedback on their thesis drafts, whereas PhD students are more likely to value their supervisors' disciplinary knowledge and help in launching their careers.

The semi-structured interviews followed a preset guide. Their purpose was to elicit first impressions about the questionnaire, explore possible resistance from students and supervisors, and gather suggestions on ways to improve the questionnaire before distributing it to a larger sample. The interviews also asked participants for their opinions regarding whether the scale would achieve its objectives of generating a valid picture of supervision quality, identifying problem situations and suggesting possible solutions, and helping students to discuss supervision quality candidly with their supervisors.

Distributing the Revised Version

The comments collected during the interviews led to a revised version of the questionnaire that was mailed to 215 students who met the study eligibility criteria listed below and were currently enrolled in eight graduate programs offered by the Faculty of Medicine and Health Sciences at the University of Sherbrooke. Seven of these programs train students in basic biomedical research while the other focuses on clinical research. As a result, 86% of the target student population receives lab-based research training that includes little coursework. To be included in the survey, students had to have been enrolled for at least one year and be able to read French. The first criterion is based on ensuring that the students have had enough opportunities to interact with their supervisors to be able to evaluate the quality of their supervision. As for the second criterion, although the University of Sherbrooke is a French-language university, a small number of its graduate students do not read French.

In order to maximize the response rate, we followed Dillman's Tailored Design Method (2000) as closely as possible. Dillman's method consists of a set of practical suggestions regarding the design of an attractive questionnaire, the ideal number of repeated mailings, and the content of each mailing. We opted for a paper questionnaire rather than an Internet-based survey for two main reasons. First, response rates to online questionnaires tend to be lower than for paper surveys; and second, online surveys have been shown to raise greater concerns about confidentiality that may discourage participation (Carini et al., 2003; Sax, Gilmartin, & Bryant, 2003).

Potential respondents were mailed a first copy of the questionnaire with a self-addressed, stamped return envelope and a personal covering letter stating the objective of the survey and underscoring the importance of their participation. A copy of the questionnaire was simultaneously e-mailed to all supervisors for their information. Seven, 14 and 21 days later, all students were reminded by e-mail to complete and return the questionnaire; another copy of the questionnaire was attached to the last reminder.

Analyzing the Items of the Revised Scale

A critical question in the analysis was how to combine the ratings for a specific item if the student had more than one supervisor, as illustrated earlier. We opted for a scoring system based on the concept of team supervision, recognizing that co-supervisors bring different skills, knowledge, and experiences to the relationship (Pole, 1998). For example, one supervisor may have content expertise, while another's strengths lie along methodological lines. One of two ratings was applied to a given item: the maximum rating when the involvement of at least one supervisor seemed sufficient (e.g., My supervisor made sure I was comfortable in my research environment) or the mean rating when we felt that all co-supervisors should meet the criterion addressed by the item (e.g., My supervisor's feedback contributes to the advancement of my project). "Does not apply" responses were ignored in the latter case.

The distribution of the responses on the different items of the questionnaire was examined using histograms and descriptive statistics. In addition, Cohen's weighted kappa and 95% confidence interval were computed to capture the variability in student assessment of the quality of the supervision provided by each of their research supervisors. A global score of satisfaction with the supervisory team was then computed by averaging the respondents' answers to the 19 G3S-SP items dealing specifically with the quality of student supervision received. Expecting few missing data, the average was computed on answered items. The global score was compared with other questionnaire items to document its validity. Lastly, the items that make up the global score were subjected to a Promax factor analysis to identify the dimensions underlying the scale. We hoped to receive at least 100 completed questionnaires as this is the minimum sample size needed to ensure the relative stability of the latent constructs that emerge from a factor analysis (Guadagnoli & Velicer, 1988; Hatcher, 1994).

RESULTS

Outcome from Focus Groups and Interviews

Eighteen individuals participated in this part of the study. A synthesis of their comments revealed some similarities and differences between the three groups of participants. In general, more similarities were found between the comments of the master's and PhD students than between the two groups of students and the supervisors.

Regarding similarities, all three groups supported implementing systematic assessment of the perceived quality of supervision, at least for the purpose of generating a global picture. Both students and supervisors thought the proposed scale was quite comprehensive. Respondents found it quick and easy to fill out, taking only about 15 minutes to complete. The questions were considered intelligible and pertinent. Respondents liked the scale's appearance, especially its uncluttered layout with adequate space for comments. Only the supervisors, however, showed any real enthusiasm about the scale; students doubted that it would help resolve conflict situations.

The difference of opinion between the students and supervisors was more marked on the question of a possible dialogue between the two parties. The students seriously doubted that the scale could help them openly discuss their dissatisfaction with their supervisors. Some argued that the risk of repercussions was too great. Expressing their grievances would be akin to shooting themselves in the foot; they preferred to suffer in silence.

The supervisors showed less scepticism about the idea that the scale could open the door to a sincere and constructive dialogue between students and supervisors. They were initially surprised that the students might feel uncomfortable expressing their dissatisfaction. It was only after being asked about their personal experiences with supervisors during their own graduate studies that they recognized the difficulty students might have in expressing themselves openly.

The greatest difference among the three groups of participants pertained to the notion of anonymity. The majority of the students, both master's and PhD, felt that the student's name should not appear on the questionnaire because, if it did, respondents would not express their dissatisfaction for fear of the consequences. The master's students were not particularly concerned about the idea of identifying their supervisors: since more of them had the same supervisor, they were less afraid of being recognized. This was not the case with the PhD students, who objected more strongly to identifying their supervisors. Some even suggested removing the sociodemographic questions, especially those regarding the respondent's age and sex, to prevent identification.

Both groups of students agreed that the supervisors should not have access to individual questionnaires, since personal details, written comments, and even handwriting would make it easy to identify respondents. The supervisors were less concerned about anonymity and had no objection to being identified. Some noted that it would be easier to interpret the results relating to them per-

sonally if they were given their students' individual questionnaires rather than a compilation of their respective evaluations. Others expressed concern about the possible uses of the data: What will be done with it? Will they be the only ones to have access to the information? Will it be given to program administrators or faculty authorities? Clearly, there was considerable disparity between the concerns of the students and those of the supervisors.

Once the interviews had been conducted and given that the next step was to explore the validity of the G3S-SP, we decided that the respondents would not be asked to identify themselves or their supervisors. This meant abandoning the objective of identifying conflict situations in order to try to resolve them. This double anonymity was emphasized in the covering letter sent with the questionnaire to avoid possible resistance from some students. Minor changes were also made to the questionnaire. This was mainly a matter of rewording some items to increase clarity. Copies of the revised version of the questionnaire are available in French and English from the corresponding author.

Measurement Properties of the Revised Scale

A total of 120 questionnaires were returned, for an overall response rate of 56%. Response rates varied considerably across the eight graduate programs, from 18% to 88%. Table 1 provides a summary of respondent characteristics and answers to the four introductory questions. Survey participants did not differ significantly from the general body of students enrolled in graduate programs at the Faculty of Medicine and Health Sciences with respect to level (M.Sc. or PhD, $p = 0.969$), sex ($p = 0.733$), or age ($p = 0.102$), the only variables for which institutional data were available.

Table 1 shows that 25 of the respondents were in their first year of study, despite having asked graduate program administrators to send us only the names of students who had been registered for more than one year. We decided to include these students in subsequent analyses, in part, because they made up a substantial portion of the sample. In addition, there was no statistically significant difference in the mean degree of satisfaction of these students compared to that of those enrolled for more than a year ($p = 0.758$).

Few data were missing for questions pertaining to the quality of supervision, with only one or two students leaving an item blank. The "Does not apply" (n/a) response appeared infrequently, except for two items concerning the dissemination of student results (8 n/a) and respect for intellectual property (15 n/a). Not surprisingly, most students who chose the n/a option were in their first or second year of studies. No items were discarded after examining the distribution of the ratings or the students' written comments. The former could have pointed to items for which a response option was disproportionately used (or not used), while the latter could have suggested the need to discard an item considered redundant.

Scores on the 19 G3S-SP items range from a mean of 2.61 out of 4 ($sd = 0.96$) for "I feel comfortable talking to my supervisor about my dissatisfaction

Table 1. Respondent characteristics and answers to the four introductory questions

Characteristics	Frequency (%)
Level	
Master's	70 (58.3)
PhD	50 (41.7)
Year of enrolment	
1 st	25 (21.0)
2 nd	50 (42.0)
3 rd	22 (18.5)
4 th or more	22 (18.5)
<i>missing</i>	1
Sex	
Female	61 (51.7)
Male	57 (48.3)
<i>Missing</i>	2
Age (in years)	
20-24	41 (34.8)
25-29	61 (51.7)
30-34	12 (10.2)
35 or more	4 (3.4)
<i>missing</i>	2
Stage	
Coursework	2 (1.7)
Research	65 (54.2)
Writing thesis	46 (38.3)
Thesis submitted	7 (5.8)
Statement ^a	
I am adequately prepared for graduate studies.	3.40 (0.61)
I have clear expectations of the program I am enrolled in.	3.21 (0.71)
I have clear expectations of my research supervisor(s).	3.38 (0.75)
I know what resources are available in the event of a disagreement with my supervisor(s).	2.62 (0.98)

^a Response options varied from 1 (completely disagree) to 4 (completely agree).

with his/her supervision" to 3.71 ($sd = 0.56$) for "My supervisor respects the intellectual property resulting from my work". Eighty-seven of the 120 respondents had only one supervisor, 27 had two, and six had three. Weighted kappa coefficients, computed for the 27 students with two supervisors, ranged from 0.15 (95% C.I. from -0.21 to 0.51) for the item concerning the supervisor's availability as compared to the student's needs to 0.81 (95% C.I. from 0.57 to 1.00) for that regarding intellectual property. Nine of the 19 kappa coefficients were below 0.50. In part, low kappas reflect our a priori hypothesis that co-supervisors have

distinct roles that complement each other, especially in the context of cross-disciplinary research. The low kappa values also follow from the asymmetrical distributions of the responses (Cicchetti & Feinstein, 1990; Feinstein & Cicchetti, 1990), most students having given ratings of 3 and 4 to their supervisors. As expected in such instances, the observed proportions of agreement were higher than corresponding indices of concordance, ranging from 48% to 90%. Twelve (36.4%) of the 33 students with more than one supervisor said that this caused them particular problems. The most common were difficulties getting co-supervisors to meet together with the student and managing differences of opinion.

The global score for satisfaction with the supervisory team, defined by averaging the 19 ratings of the G3S-SP, had a mean of 3.39 with a standard deviation of 0.46 (median = 3.47). Mean satisfaction scores differed significantly across graduate programs, from 3.14 ± 0.69 to 3.68 ± 0.24 ($p = 0.036$), an indication of the discriminant validity of the scale. It also differed between the 12 students who expressed difficulties with having more than one supervisor (3.37 ± 0.33) and those who did not (3.52 ± 0.36). The difference was not statistically significant ($p = 0.212$), in part, because only 33 students had more than one supervisor. On the other hand, the global satisfaction score was highly correlated with two questionnaire items asking students for an overall assessment of supervision quality, specifically: "All things considered, would you recommend this research supervisor to another student? *Yes, unconditionally; Yes, but with reservations; No*" (Spearman's $\rho = -0.56$, $p < 0.001$) and "All things considered, how would you evaluate the supervision you get from your supervisor? From *completely unsatisfactory to completely satisfactory*" ($\rho = 0.74$, $p < 0.001$). When asked for suggestions on how their supervisors could improve the quality of their supervision, 46 students suggested increasing availability, while 10 others suggested improving their interpersonal skills. Other suggestions were made by fewer than eight students. Some suggestions were quite blunt, such as that of one student who wrote: "Change job!"

Very few students expressed the desire to meet with graduate program administrators ($n = 6$) or their student representatives ($n = 9$) to discuss supervision quality. On average, the global satisfaction score of those who answered "yes" to one of these two questions (3.16 ± 0.69 , $n = 10$) was lower than that of those students who did not feel the need to discuss their situation with others (3.41 ± 0.43 , $n = 110$). The difference, however, was not statistically significant ($p = 0.330$), in part, because of the small number of students in the "yes" category. Still fewer said they took advantage of our invitation to complete the questionnaire to discuss the delicate issue of supervision with their supervisors ($n = 7$). When asked why not, most students replied that they simply did not feel the need, being generally satisfied with the quality of supervision. Many of the remainder felt uncomfortable discussing this subject with their supervisors, sometimes out of fear of reprisals.

The two items with higher rates of n/a responses were excluded from the factor analysis conducted to explore the underlying structure of the G3S-SP. As

a result, the analysis was based on a reduced sample of 105 respondents without any missing data on the remaining 17 items of the scale. The analysis revealed a two-factor structure, with eigenvalues of 5.57 and 1.13, respectively, which accounted for 84% of the variance. The correlation between the two factors was 0.50. Table 2 presents the loading of each item on these factors and the item-factor correlation coefficients. Factor 1 accounts for 70% of the variance and includes nine items (Cronbach's $\alpha = 0.88$) that mainly reflect the supervisor's help with designing and conducting the student's research project. The other eight items classified under Factor 2 ($\alpha = 0.76$) relate to the relationship aspect. Although classified under Factor 1, the item "Since my registration, my supervisor and I have discussed how we will collaborate" refers more to the student's relationship with his/her supervisors than the latter's involvement in the student's research project. This suggests that this item might equally well be classified under Factor 2, especially since it had a loading of 0.32 on that factor.

At the end of the questionnaire, respondents were asked if there were any aspects of their supervision that were not covered by the scale and should be added. The most frequent answer concerned the student's financial support. Inadequate financial support has been shown to decrease student possibilities to successfully complete graduate studies in a relatively short time (Ferrer de Valero, 2001). In one item, students were asked whether the information provided by their supervisors about financial resources was adequate; no item specifically asked students if they were having financial difficulties. A question on this topic should be added. Despite this omission, when invited to rate the usefulness of the scale in describing the quality of their supervision, 90% of respondents answered *quite useful* or *very useful*. Of greater interest, perhaps, are the students' comments regarding the scale. Some mentioned that it raised their awareness of aspects that should be discussed with their supervisors shortly after having engaged in the relationship. Others noted that even if the scale generated an accurate picture of the quality of supervision, it could not be used to identify problem situations because of the double anonymity. Thus no help could be directed to those students who might need it. Others pointed out that it is very difficult to change supervisory practices. In their opinion, removing anonymity would not ensure that conflict situations would ultimately be resolved.

DISCUSSION

This study has provided a tool for measuring health sciences graduate students' perceptions of the quality of their supervision, together with some evidence of its validity. The sample was restricted to research trainees from a single university, all of whom were enrolled in a North American health sciences faculty. It remains to be seen whether the G3S-SP equally applies to students from other graduate schools as well as to those trained in other areas (the arts and humanities, for example). In the sciences, graduate student research projects are typically extensions of the supervisor's work. This situation is less frequent in non-science fields which could affect the validity of some items of

Table 2. Results of the Promax factor analysis conducted on 17 items of the G3S-SP^a

Factor and Item ^b	Loading on Factor 1	Loading on Factor 2	Item-factor Correlation
Factor 1: Supervisor's involvement in the student's research project			
My supervisor helped me structure the steps in my research.	0.88		0.71
My supervisor helped me choose and clarify my research subject.	0.77		0.72
My supervisor helped me set the limits of my research project.	0.70		0.66
The speed of feedback from my supervisor in relation to my project is: (completely inadequate to completely adequate)	0.59		0.73
The availability of my supervisor as compared to my needs is: (completely inadequate to completely adequate)	0.59		0.61
My supervisor's feedback contributes to the advancement of my project.	0.57		0.69
My supervisor shows enthusiasm for my project.	0.54		0.61
My supervisor has good disciplinary knowledge in my research area.	0.45		0.49
Since my registration, my supervisor and I have discussed how we will collaborate.	0.40		0.51
Factor 2: Interpersonal relations			
I have a good professional relationship with my supervisor.		0.77	0.55
I have a good interpersonal relationship with my supervisor.		0.75	0.59
I feel comfortable talking to my supervisor about my dissatisfaction with his/her supervision		0.56	0.52
My supervisor made sure I was comfortable in my research environment (physical and social).		0.41	0.57
My supervisor's requirements about the work I have to deliver are: (completely unreasonable to completely reasonable)		0.39	0.43
The material resources (work area, computer, technical support, etc.) to which my supervisor has given me access are: (completely inadequate to completely adequate)		0.38	0.40
The information my supervisor gave me to introduce me to the scientific community (meetings, research teams, etc.) is: (completely inadequate to completely adequate)		0.30	0.28
The information my supervisor gave me about financial resources (scholarships, research assistantships, stipend from his/her research funds, etc.) is: (completely inadequate to completely adequate)		0.26	0.42

^a Excluding the two items that had higher rates of "Does not apply" responses.

^b The response scale varied from 1 (completely disagree) to 4 (completely agree), except where indicated.

the scale and its underlying structure. The amount of coursework required from students may also impact their satisfaction with graduate supervision. Additionally, expectations of the supervisory relationship may differ across cultures and countries. These issues should be examined in future studies, using larger and more diverse samples that would provide greater statistical power and more generalizable findings. Such samples would also facilitate identifying student attributes that influence satisfaction with graduate supervision, such as age, sex, academic discipline, and stage of advancement (Rose, 2005).

In this study, we accounted for joint supervision by asking students to provide separate ratings for each of their supervisors. Ratings, however, were restricted to supervisors formally involved in supervision. In the health sciences, many students conduct their research as part of a wider group whose members are all engaged in the same research program. Such groups often include other faculty members, postdoctoral fellows, technicians, and other graduate students (Pole, 1998). Collectively, the group provides important support to the student, which our scale does not currently capture. Perhaps future efforts should aim at incorporating this extra support in the assessment of students' perception of the quality of their supervision.

Following feedback from participants in the focus groups, we decided not to ask students to identify themselves, which precludes linking individual evaluations to particular supervisors. Perhaps students would be more willing to identify their supervisors if they were surveyed after graduation, as is done in Australia (ACER, 2000). March, Rowe, and Martin (2002, p. 340) have noted, however, that students asked to rate the quality of their supervision, knowing that ratings would be returned to their supervisors, would have a serious conflict of interest. Unlike evaluations of classroom teaching, any one supervisor is unlikely to have many students completing their degree in a given year, which would preclude any effective guarantee of the confidentiality of their responses. Yet students are likely to be dependent on supervisors for letters of reference to prospective employers for at least the early part of their subsequent career. Hence scales like the G3S-SP are more likely useful to portray supervision quality in a given program, as perceived by graduate students, and monitor change over time.

A critical issue then becomes whether the picture drawn from students' assessments is valid. Our response rate was higher than that observed in other surveys of graduate students (ACER, 2000; March, Rowe, & Martin, 2002), perhaps because of the perceived relevance of the survey to the students' current lives (Sax, Gilmartin, & Bryant, 2003). Moreover, the students who returned the questionnaire were representative of those who were invited to do so, at least with respect to demographics and program level. They may, however, differ to some extent from those who chose not to participate in the study. In particular, the latter may be somewhat less satisfied with their supervision than the observed data suggest. Many factors likely enter into the decision not to participate in a survey. It is nonetheless possible that dissatisfied students were reluctant to give negative ratings to their supervisors, upon whom they continue

to depend. When the total population size is small, as in this study, anonymity may not be a sufficient guarantee for those who fear repercussions.

While refusal bias cannot be ruled out entirely, a number of written comments suggest that some dissatisfied students did elect to fill out the questionnaire. Although the data are somewhat skewed towards the high satisfaction end of the scale, a small yet significant proportion of the respondents expressed dissatisfaction through their ratings. For example, when asked whether they would recommend their supervisors to other students, 8% answered that they would not in the case of at least one supervisor. An additional 48% replied that they would, but with reservations. Similarly, when invited to rate their overall level of satisfaction, 18% indicated that they were completely or somewhat dissatisfied with at least one of their supervisors.

Despite being derived from a sample that likely includes some less satisfied students, the observed data reflect a generally favourable attitude on the part of health sciences research trainees towards their supervisors. High scores may reflect low expectations or limited frames of reference rather than true satisfaction. In addition, what students view as *adequate* or *reasonable* may be influenced by a number of factors, including culture, academic discipline and desire to be trained by a renowned researcher. High satisfaction scores may also be the result of framing all items in the same direction, with higher ratings always indicating greater satisfaction. Perhaps some items should have been framed in negative terms and their scale inverted when deriving the global satisfaction score. The statistically significant correlation between the global score and the overall assessment of supervision quality, whose scale was reversed, suggests that stating all items positively is unlikely to have invalidated the results. On the other hand, the fact that this correlation was far from unity (Spearman's $\rho = -0.56$) supports the use of multiple items for assessing the quality of research trainee supervision. In addition to making students aware of important dimensions of graduate supervision, ratings of multiple items provide valuable information to program administrators regarding areas that students perceive as needing improvement.

A quality score derived from averaging a student's ratings over multiple items is also more reliable than a single overall satisfaction item. The test-retest reliability of the students' ratings was not examined in this study and should be investigated in the future. Because the relationship between students and supervisors may change over time (Pole, 1998), the interval between repeated administrations of the G3S-SP should be short enough to ensure that the construct being measured is stable. Under this condition, the reliability of the G3S-SP should be quite high, given the equivalence between Cronbach's alpha and the intraclass correlation coefficient (Bravo & Potvin, 1991). The same holds true for the two underlying dimensions of the scale that emerged from factor analysis. Given the exploratory nature of this analysis, its results should be interpreted with caution and the two dimensions cross-validated in other samples. Meanwhile, the factor analysis suggests that health sciences gradu-

ate student perceptions of the quality of their supervision rests on two sets of considerations: one reflecting practical assistance with the research project; the other representing student-supervisor relationships. Practical guidance and relationship have also been found to be essential functions of *mentoring* in the academic setting, a concept that is related to but distinct from that of supervising research trainees (Rose, 2003, 2005). The construct validity of the G3S-SP could also be further investigated by testing hypotheses relating the global satisfaction score to other constructs with which it should theoretically be linked. For example, one could hypothesize that perceived supervision quality should be correlated, at least moderately, with degree attainment and time-to-degree.

In conclusion, the study results suggest that the G3S-SP could be a useful tool in generating a reliable and valid picture of the quality of health sciences research trainee supervision in a given graduate program, as perceived by students. As for course evaluations, the picture could take the form of a "program report card" that includes descriptive statistics on the global satisfaction score (measures of central tendency and dispersion), identifies areas of marked dissatisfaction (and satisfaction), and summarizes students' written comments. It remains to be seen whether the scale would be responsive, that is, sensitive to changes induced by local efforts to resolve the problems identified. If respondents do not provide their names, the scale cannot be used to identify students and/or supervisors to whom these efforts should be directed. A qualitative rather than quantitative approach, in which students would describe the quality of their supervision to a neutral person, could prove more effective in identifying conflict situations and taking the necessary corrective action. In this context, the G3S-SP could be used to structure the discussion and direct student attention to discrete components of their supervision.

ACKNOWLEDGEMENTS

This work was conducted with the financial support of the University of Sherbrooke. The authors thank the members of the Clinical Sciences Program Committee for their comments on the study protocol. They also thank all the students and supervisors who took part in the development and evaluation of the G3S-SP, as well as the directors of the eight graduate programs of the Faculty of Medicine and Health Sciences for allowing the questionnaire to be distributed to the students. We are also grateful to Professor Jean Nicolas of the University of Sherbrooke for enlightening discussions about graduate supervision. Lastly, we thank the editor and three anonymous reviewers for their helpful comments on the paper.

REFERENCES

- Adam, K.A. (2002). *What colleges and universities want in new faculty*. Association of American Colleges and Universities.
- Anderson, E.M., & Shannon, A.L. (1988). Toward a conceptualization of mentoring. *Journal of Teacher Education*, 39(1), 38-42.
- Anderson, M.S., & Swazey, J.P. (1998). Reflections on the graduate student experience: An overview. *New Directions for Higher Education*, 26(1), 3-13.
- Association for Support of Graduate Students. (1993). *Survey of thesis difficulties*. Retrieved October 29, 2004 from the World Wide Web at http://www.asgs.org/Annl_Svy.htm.
- Australian Council for Education Research (ACER). (2000). *Evaluation and validation of the trial Postgraduate Research Experience Questionnaires*. Retrieved August 17, 2007 from the World Wide Web at <http://www.dest.gov.au/archive/highered/eippubs/eip99-10/preq.pdf>
- Bravo, G., & Potvin, L. (1991). Estimating the reliability of continuous measures with Cronbach's alpha or the intraclass correlation coefficient: Toward the integration of two traditions. *Journal of Clinical Epidemiology*, 44(4/5), 381-390.
- Brown, G., & Atkins, M. (1988). *Effective teaching in higher education*. London & New York: Methuen.
- Burlew, L.D. (1991). Multiple mentor model: A conceptual framework. *Journal of Career Development*, 17(3), 213-221.
- Canadian Association for Graduate Studies. (1997). *Time to completion and cohort completion percentages by discipline: A consolidation of data for the 1985-1988 cohorts*. St. John's, NL: Author.
- Canadian Association for Graduate Studies. (2004). *The completion of graduate studies in Canadian universities. Report & Recommendations*. Ottawa, ON: Author.
- Carini, R.M., Hayek, J.C., Kuh, G.D., Kennedy, J.M., & Ouimet, J.A. (2003). College student responses to web and paper surveys: Does mode matter? *Research in Higher Education*, 44(1), 1-19.
- Cicchetti, D.V., & Feinstein, A.R. (1990). High agreement but low kappa II: Resolving the paradoxes. *Journal of Clinical Epidemiology*, 43(6), 551-558.
- Council of Graduate Schools in the United States. (1990). *Research student and supervisor, an approach to good supervisory practice*. Washington, DC: Author.
- Denecke, D. (2005). Ph.D. Completion project: Preliminary results from baseline data. *Communicator*, XXXVIII (9), 1-4.

Dillman, D.S. (2000). *Mail and internet surveys: The tailored design method*. New York: John Wiley & Sons.

Donald, J.G., Saroyan, A., & Denison, D.B. (1995). Graduate student supervision policies and procedures: A case study of issues and factors affecting graduate studies. *Canadian Journal of Higher Education*, 25(3), 71-92.

Elgar, F.J. (2003). *PhD degree completion in Canadian universities. Final report*. Halifax, NS.

ESRC. (1994). *Studentship Handbook*. Swindon: Author.

Farr, M. (2002, November). 'Til degree do we part. *University Affairs*, 10-13.

Feinstein, A.R., & Cicchetti, D.V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, 43(6), 543-549.

Ferrer de Valero, Y. (2001). Departmental factors affecting time-to-degree and completion rates of doctoral students at one land-grant research institution. *Journal of Higher Education*, 72(3), 341-367.

Green, S.G., & Bauer, T.N. (1995). Supervisory mentoring by advisers: Relationships with doctoral student potential, productivity, and commitment. *Personnel Psychology*, 48(3), 537-561.

Guadagnoli, E., & Velicer, W.F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 103(2), 265-275.

Hahs, D.L. (1998, November). Creating "good" graduate students: A model for success. Paper presented at the *Annual Meeting of the Mid-South Educational Research Association*, New Orleans.

Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling*. Cary, NC: SAS Institute Inc.

Henderson, P.H., Clark, J.E., & Reynolds, M.A. (1996). *Summary Report 1995: Doctoral recipients from United States universities*. Washington, DC: National Academy Press.

Higher Education Funding Council for England. (2005). *PhD research degrees: Entry and completion*. Retrieved August 17, 2007 from the World Wide Web at http://www.hefce.ac.uk/pubs/hefce/2005/05_02/

Hinchey, P., & Kimmel, J. (2000). *The graduate grind, a critical look at graduate education*. New York, NY: Falmer Press.

Hockey, J. (1995). Getting too close: A problem and possible solution in the social science PhD supervision. *British Journal of Guidance & Counselling*, 23(2), 199-210.

Holdaway, E.A. (1994). Organisation and administration of graduate studies in Canadian universities. *Canadian Journal of Higher Education*, 24(1), 1-29.

Holdaway, E.A. (1996). Current issues in graduate education. *Journal of Higher Education Policy & Management*, 18(1), 59-74.

Jacobi, M. (1991). Mentoring and undergraduate academic success: A literature review. *Review of Educational Research*, 61(4), 505-532.

Kelly, E. (1998). *Strengthening graduate education in science and engineering*. National Science Board Keynote Remarks. Arlington, VA: National Institute for Science Education, Graduate Education Forum.

Kerlin, S. (1995). *Pursuit of the PhD: Survival of the fittest, or is it time for a new approach?* Retrieved October 29, 2004 from the World Wide Web at <http://epaa.asu.edu/epaa/v3n16.html>.

Knowles, M. (1975). *Self-directed learning*. New York: Associated Press.

Lapidus, J.B. (1997). Issues and themes in postgraduate education in the United States: Beyond the first degree. In R. G. Burgess (Ed.), *Society for research into higher education* (pp. 22-39). San Francisco, CA: Jossey Bass.

Lovitts, B.E. (2001). *Leaving the ivory tower. The causes and consequences of departure from doctoral study*. New York: Rowman & Littlefield Publishers Inc.

March, H.W., & Roche, L.A. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal*, 30(1), 217-251.

March, H.W., & Roche, L.A. (1997). Making students' evaluation of teaching effectiveness effective. *American Psychologist*, 52(11), 1187-1197.

Marsh, H.W., Rowe, K.J., & Martin, A. (2002). PhD students' evaluations of research supervision: Issues, complexities, and challenges in a nationwide Australian experiment in benchmarking universities. *Journal of Higher Education*, 73(3), 312-348.

McKeachie, W.J. (1997). Student ratings: The validity of use. *American Psychologist*, 52(11), 1218-1225.

National Research Council. (1995). Increasing graduate student retention and degree attainment. In L. Baird (ed.), *New directions for institutional research*, No. 80. San Francisco, CA: Jossey Bass.

Nerad, M., & Cerny, J. (1999). Postdoctoral patterns, career advancement, and problems. *Science*, 285(5433), 1533-1535.

Nyquist, J.D. (2002). The PhD: A tapestry of change for the 21st century. *Change*, pp. 12-20.

Pearson, M. (1996). Professionalizing PhD education to enhance the quality of the student experience. *Higher Education*, 32(3), 303-320.

Pole, C. (1998). Joint supervision and the PhD: Safety net or panacea? *Assessment and Evaluation in Higher Education*, 23(3), 259-272.

Pole, C., Sprokkereef, A., & Burgess, R. (1997). Supervision of doctoral students in the natural sciences: expectations and experiences. *Assessment and Evaluation in Higher Education*, 22(1), 49-64.

Ramos, M.G. (1994). *Understanding the ABC (all but dissertation) doctoral candidate: A phenomenological approach*. Unpublished doctoral dissertation, University of Kansas.

Roberts, G.C., & Sprague, R.L. (1995). To compete or to educate? Mentoring and the research climate. *Professional Ethics Report*, VIII(4), 6-7.

Rose, G.L. (2003). Enhancement of mentor selection using the Ideal Mentor Scale. *Research in Higher Education*, 44(4), 473-494.

Rose, G.L. (2005). Group differences in graduate students' concepts of the ideal mentor. *Research in Higher Education*, 46(1), 53-80.

Sax, L.J., Gilmartin, S.K., & Bryant, A.N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education*, 44(4), 409-432.

SERC. (1992). *Research student and supervisor: An approach to good supervisor practice*. Swindon: Author.

Tinto, V. (1993). *Leaving College: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago, IL: University of Chicago Press.

Tluczek, J.L. (1995). *Obstacles and attitudes affecting graduate persistence in completing the doctoral dissertation*. Unpublished doctoral dissertation, Wayne State University.

Winston, R.B., & Polkosnik, M.C. (1984). Advising graduate and professional school students. In R. B. Winston, T. K. Miller, S. C. Ender, and T. J. Grites (Eds.), *Developmental academic advising: Addressing students educational career and personal needs*. San Francisco, CA: Jossey Bass.

CONTACT INFORMATION

Professor Gina Bravo
Department of Community Health Sciences
Faculty of Medicine and Health Sciences
University of Sherbrooke
3001 12th Avenue North
Sherbrooke, QC
Canada, J1H 5N4.
Telephone: (819) 564-5361; Fax: (819) 564-5397
E-mail: Gina.Bravo@USherbrooke.ca