National Capital Chapter
Canadian Evaluation Society

Evaluation Case Competition

Case for the Preliminary Round

February 1999

"DOCTORAL RESEARCH AWARDS"

This material, a combination of actual and hypothetical documents, has been assembled for educational purposes only.

About the Canadian Evaluation Society and the Evaluation Case Competition

The Canadian Evaluation Society, linked to similar societies around the world, is a non-profit organization dedicated to improving the quality of evaluations and their use in decision-making. One facet of its work is raising awareness among students about the importance of solid evaluative information to the successful administration of firms, government initiatives, educational programs and non-profit organizations - and it here that the case competition plays a role. The competition provides university students with an intense, hands-on experience in assessing and condensing information related to an evaluation then presenting observations and recommendations to a decision-making body.

Here is how the Competition works. The National Capital Chapter of the Evaluation Society invites universities to assemble teams of three to five students. In a first round of the competition, teams have five hours to examine an evaluation case file (available in both official languages) then submit a written analysis by fax for review by a panel of judges. The three top-rated teams are invited to Ottawa for a final round. There they tackle a new case then present their analyses to the judges in front of a live audience.   The winning team takes possession of the crystal competition trophy for a year and is awarded prizes from a fund established by sponsors of the Competition. Information on members of the winning team is circulated within the Canadian evaluation community to give them visibility with potential future employers.

More on the Canadian Evaluation Society at:

http://www.unites.uqam.ca/ces/ces-sce.html

Rules for the Preliminary Round of Competition

There must be at least three and no more than five members to a team.

All team members must be registered in a program of university studies (undergraduate or graduate, full or part time). Team members may be from any academic discipline.

Teams may be coached prior to the competition but coaches must not communicate with their teams once they (the teams) have received the case.

In preparing their submissions, teams are at liberty to explore any public information source such as would be accessible by a management consulting group. For example, they may consult books or articles, search libraries, use the Internet, and so forth. Team members are free to leave the work-site and take refreshment as they wish but they may communicate only within the team.

Teams have five hours to prepare their submission. Coaches will deliver the case materials to them at the start of this period. Submissions must be faxed to the competition organizers at the end of the five hours. **Send the fax to** Celine Pinsent at **(613) 235-9592.** Celine's telephone number is (613) 282-9069 or (613) 230-5577 (operator 0).

In addition to sending the submission by fax, teams with e-mail access are invited to submit an electronic version of their work in MSWord, WordPerfect or PowerPoint. Teams should save the document in an early version of the software as some judges may not have the more recent versions. The file may be attached to an e-mail message addressed to cpinsent@ggi.ca with a copy to rockhurst@sprint.ca. **Note that e-mail submission is not obligatory** but would help organizers explore its use for the Year 2000 Competition.

Submissions should be concise, and 1500 words is recommended as a targeted maximum. They may be in paragraph or point form, with or without graphics.

Judges are not to know the identity of the teams. Teams should reveal their true identity only on a cover page which will be removed before the submission is passed to the judges. However, they may use a non-revealing code name (e.g., XYZ Consulting, or, Students from Course ABC 123) throughout the report if they wish.

Criteria

The following table is intended to indicate to teams and judges the basis of assessment. However, the uniqueness of each case necessitates flexibility in the evaluation process.

Assessment Area	Criteria	Weight (%)
Content	Relevance of material selected for presentation	
	• relevance to issues	5
	• relevance to findings	5
	• relevance to conclusions	10
	Evidence presented	
	• support for points made	5
	• awareness of limits of evidence	5
Accessibility	Ease of absorption of information	
	• clarity of points made	10
	• logical linkage between points (ideas flow)	5
	• organization of points	10
Impact	Strength of argument	
	• cohesiveness	5
	• persuasiveness	5
	• memorableness	5
	Impetus for decision-making	
	• rationality of recommendations	15
Overall	Overall impression of the presentation	15
		100

The Scenario

You and your fellow students are taking a course entitled *Conducting and Using Research*, taught by Professor Benito Karma. The class meets for five hours every Saturday, the only time that fits with everyone's class schedule.

On February 6, 1999, when you enter the seminar room you find the following message taped to the blackboard. It is written in the distinctive script font that Professor Karma has set as the default on his computer.

Dear Students,

I must attend to an urgent family matter this weekend.

I would be very grateful if you would use class time to respond to the attached request. You will not only be making an important intellectual contribution to the Medical Research Council but should also benefit from the chance to put into practice some of the general principles we have been discussing in the course so far.

Please fax your views on the document to Dr. Jacinthe Vale-Chapeau and offer my apologies to her for not being able to reply personally.

Many thanks,

MRCRM

Medical Research   Conseil de recherches
Council of Canada   médicales du Canada

February 3, 1999

Dr. Benito Karma
University of Cascaseco
Allacrosse, Canada
I4U 4T5

Dear Dr Karma,

I enjoyed very much meeting you at the Fellowships Committee review session last October. Our lunch time discussion about evaluation was especially memorable. As explained over the phone yesterday, I would like to take up your offer to serve as an external critic of the Council's evaluative activities.

At the November 1998 meeting of the Council, board members enquired about ongoing reviews of those MRC programs that focus on training and developing the next generation of health researchers. As staff officer responsible for research personnel programs, I was asked to table at the March 1999 meeting of Council all recent evaluative studies of our training and development programs.

Attached for your consideration is a report on the assessment process for a new program, Doctoral Research Awards (DRA), that was implemented in the 1997-98 competitions cycle. The study of the DRA assessment process was conducted by a summer student during the period May to September 1998 using data that had been collected for the review of applications and a survey of members of the DRA selection committee. While notionally I was the project supervisor, I was unable to spend as much time on the study as I would have liked. As you know, the Council is heavily involved in designing a new approach to the support of health research in Canada and, as a result, time for evaluation has been in short supply.

From my perspective the summer student seemed to have a good grasp of the issues and brought to the study a strong background in statistical methods. Qualitative analysis of the survey results seems reasonable. However it has

been many years since I opened textbooks in these areas and I do not feel competent to judge the quality of the work.

Monday, February 8 is the last day for placing items on the agenda for the next meeting of the Council. I will be working on my submissions to the agenda on the weekend and would be most grateful if you would send me your review of the student's study **by Saturday afternoon**. I will be working at home, so you could fax your comments to me there.

The following questions are provided just to give you a sense of some of the areas in which I am interested. Please do not feel it necessary to structure your review around them.

- What are the key points about this study that you would communicate to the members of Council?

- Does the study appear valid? Is it likely that if someone repeated the methodology that the same observations would emerge?

- What are the strengths and weaknesses of the study? How could it be improved?

- Assuming that at least some of the observations are strong enough to support recommendations, what would you suggest be done to improve the DRA assessment process.

Thanks in advance for any feedback that you can provide. You can be sure that your comments, negative or positive, will be welcome.

Yours sincerely,

Jacinthe Vale-Chapeau, PhD
Head
Personnel Support Programs

MRCRM

Medical Research   Conseil de recherches
Council of Canada   médicales du Canada

Study of the System for Selecting

MRC Doctoral Research Awards

Edited Technical Reports

Section A:

Analysis of Data from the Assessment of DRA

Applications

Section B:

Survey of DRA Committee Members

January 1999

Part A

Analysis of Data

from the Assessment of

DRA Applications

INTRODUCTION

In 1997, the Medical Research Council of Canada (MRC) developed and implemented the Doctoral Research Award (DRA). The objective of the DRA is to provide both national recognition and financial support to doctoral students in the health sciences who show high potential for future research achievement and productivity. The DRA provides an annual stipend of $19,030 to awardees, for a maximum of three years, and an annual $500 research and travel allowance. For the first DRA competition, the MRC received 273 eligible applications of which it approved 105. The award rate in the competition was thus 38.5%.

> *This study aims to examine:*
>
> - fairness of the DRA selection criteria and review system
>
> - inter-rater reliability
>
> - inter-relationships among criteria

The primary objective of this report is to provide MRC program managers and committee members with feedback on:

- fairness of the DRA selection criteria and review system
- internal consistency and inter-rater reliability of the DRA rating form
- relative importance of, and a conceptual grouping of, the DRA selection criteria

Overview of the DRA Assessment Process

Candidates submit to the council an application package that includes: reports from three sponsors on their personal suitability for

1

an award; information on their academic and research achievements; and data on the environment in which the doctoral research will take place.

The council assigns each application to two members of the DRA Selection Committee for assessment. Committee members are provided with rating forms (Appendix A) and guidelines for their use (Appendix B). The textbox, on the following page indicates the variables considered and the weight that each carries in the overall score.

Criteria considered in the assessment of DRA applications			
Focus of the Criteria	Criteria Considered	Weight in Overall Score	Rated by
Applicant	• Critical thinking • Independence • Perseverance • Originality • Organizational skills • Interest in discovery • Research ability	5.6 5.6 5.6 5.6 5.6 5.6 5.6	Applicant`s three sponsors
	Adjustment score	5.6	Two committee members
	• Publication activity • Other research activity • Undergraduate academic performance • Graduate performance	8.3 8.3 8.3 8.3	
Environment of doctoral research training	• Scientific activity • Research resources • Training record of the proposed doctoral supervisor • Training program for the candidate	5.6 5.6 5.6 5.6	

The first seven selection criteria are assessed by the applicant`s three sponsors who must use a 4-point scale and written comment

(See Appendix C, Sponsor's Rating Form). The sponsor`s scores on these criteria are entered into a database as soon as the application is received. The sponsors' scores are then pre-printed on the rating form that is provided to committee members. Committee members may provide an adjustment score ranging from -4 to +4, if they feel that the sponsor`s ratings do not correspond to the written comments. Committee members assess the remaining eight criteria (publication activity... to ... training program) using data provided by the candidate and the doctoral research supervisor.

Committee members send their completed rating form to the MRC where scores are entered into the application database. The two scores for each application are averaged and all applications are then sorted in descending order. A tentative cut-off line is drawn in the ranked list on the basis of funds available and MRC expectations of excellence.

Once the tentative cut-off point is established, the MRC identifies applications in which there was a very large spread between the two reviewers' ratings which, if narrowed, might move the application across the cut-off line. (For the first competition, a very large spread was defined as a difference between the scores of two reviewers that was more than one standard deviation from the mean spread for the entire competition.) For those applications, a third review was obtained and a new mean score calculated. Applications were ranked again and a new cut-off was drawn.

This Study

This study applied a variety of statistical tests to the database of applicants, reviewers and ratings developed for administration of the DRA Competition to help assess issues of fairness and reliability.

RESULTS

1. DESCRIPTIVE DATA

Candidates

For the first Doctoral Research Award competition (1997-98), the Medical Research Council of Canada received a total of 273 eligible applications, 147 (53.8%) from males and 126 (46.2%) from females. With regard to language preference of the DRA candidate, 212 (77.7%) preferred to correspond in English and 61 (22.3%) preferred French correspondence. The provincial distribution of the DRA candidates by university of proposed doctoral study was as follows:

- Quebec = 134 (49.1%)
- Ontario = 83 (30.4%)
- British Columbia = 25 (9.2%)
- Alberta = 18 (6.6%)
- Saskatchewan = 4 (1.5%)
- Nova Scotia = 4 (1.5%)
- Manitoba = 3 (1.1%)
- New Brunswick = 1 (0.4%)
- Newfoundland = 1 (0.4%)

No applications were received from Prince Edward Island.

Committee Members

The Doctoral Research Award Committee consists of a chairperson, 5 core committee members and 29 committee

Applicants, Award Rates and Committee Members

	Women	Men	Total
DRA applicants	126	147	273
Award rate	39.7%	37.4%	38.5%
DRA committee members	13	21	34

members. The core committee provides consultation on program policy and oversees the review process.   Among the 34 committee members who reviewed applications, there were 13 (38.2%) women and 21 (61.8%) men.

All but three (8.8%) of the 34 committee members indicated English as their preferred language for correspondence. On average, DRA committee members reviewed 16 applications with the number of applications reviewed ranging from 12 to 20.

Selection Criteria

The average application score was 58.04 ($\underline{SD}$ = 4.95) out of a maximum 72 with scores ranging from 43.5 to 70.2.

Award Rates

To receive an award in the first competition, a DRA candidate needed to obtain an overall rating of 59.8 or higher (the equivalent to a qualitative rating of "excellent" or better). The award rate was 38.5%.

2.   EXAMINING FAIRNESS

2.1   General Examination of Fairness

2.1.1 Gender

Overall, there was no significant difference in award rates between male (37.4%) and female (39.7%) candidates ($\chi^2(1) = .147$, ns). Moreover, a series of t-tests indicated no significant mean difference between female and male candidates for DRA total score or for those eight selection criteria which were assessed by the committee members.

However, a gender difference was found for scores on four of the eight selection criteria assessed by the DRA candidates' sponsors. There was a significant mean difference between female and male candidates on scores of perseverance, organizational skills, originality and interest in discovery.

Gender Differences in
Four Sponsor-Assessed Selection Criteria

	Female Applicants	Male Applicants
Perseverance	Higher 3.9	3.8
Organizational skills	Higher 3.8	3.7
Originality	3.4	Higher 3.5
Interest in discovery	3.8 (3.78)	Higher 3.8 (3.84)

T-tests indicated that the differences were statistically significant **but** measures of effect size indicated that gender accounted for only a small percentage of the variance in the criteria.

- In comparison to sponsor's scores of **perseverance** for male DRA candidates ($\underline{M} = 3.79$, $\underline{SD} = 0.25$), scores for female candidates ($\underline{M} = 3.86$, $\underline{SD} = 0.19$) were   significantly higher ($\underline{t}(271) = 2.73$, $\underline{p} < .01$; $\underline{\omega}^2 = .027$). Gender of applicant accounted for 2.7% of the variability in perseverance.

- Also, female DRA candidates ($\underline{M} = 3.81$, $\underline{SD} = 0.24$) tended to receive significantly higher scores from sponsors for **organizational skills** than did male ($\underline{M} = 3.72$, $\underline{SD} = 0.28$) candidates ($\underline{t}(271) = 2.76$, $\underline{p} < .01$; $\underline{\omega}^2 = .027$). Gender of applicant explained 2.7% of the variability in organizational skills.

- Sponsors' scores on **originality** for male candidates ($\underline{M}$ = 3.48, $\underline{SD}$ = 0.34) were significantly higher when compared to female ($\underline{M}$ = 3.39, $\underline{SD}$ = 0.36) candidates ($\underline{t}$(271) = 2.13, $\underline{p}$ < .05; $\underline{\omega}^2$ = .016). Gender of applicant accounted for 1.6% of the   variability in originality.

- Finally, male DRA candidates ($\underline{M}$ = 3.84, $\underline{SD}$ = 0.23) received higher scores from sponsors on **interest in discovery** when compared to female ($\underline{M}$ = 3.78, $\underline{SD}$ = 0.27) candidates ($\underline{t}$(271) = 2.12, $\underline{p}$ < .05; $\underline{\omega}^2$ = .016). Gender of applicant explained 1.6% of the variability in interest in discovery.

Although the gender differences were statistically significant, gender explained only a small percentage of the variability in those criteria. That is, the size of the gender effect was small. This finding suggested that significant differences detected by t-tests may to some extent be an artefact of the large sample size and may not indicate a meaningful difference. It is worth noting that when scores for all seven of the sponsor-assessed criteria are considered as a block, there was no statistically significant difference in mean scores between female ($\underline{M}$ = 25.83, $\underline{SD}$ = 1.37) and male ($\underline{M}$ = 25.85, $\underline{SD}$ = 1.28) candidates ($\underline{t}$(271) = .158, ns).

2.1.2 Language

Language was assessed in terms of both language preference of the candidate and language of the application. The correlation between language preference and language of the application was .64 ($\underline{p}$ < .001).   Overall, there was no significant difference in award rates between English language (38.7%) and French language (37.7%) DRA candidates (χ^2(1) = .019, ns). Likewise, applications written in French (32.3%) had award rates comparable to applications written in English (39.3%; χ^2(1) = .569, ns).

Mean scores for English language and French language DRA candidates overall or on individual selection criteria were not

statistically different.   Likewise, there were no significant mean differences between applications written in English and applications written in French overall and on the selection criteria.

2.1.3 Region

To assess whether or not a regional bias might have operated within the DRA competition, a chi-square test for goodness of fit was attempted using Canadian provinces as the basis for comparisons. However, 55.6% of the cells in the table of province by award rate had expected frequencies of < 5. Provinces were therefore collapsed into the following three Canadian regions: Eastern, Central or Western. However, even after this reclassification of categories, Eastern Canada had too few cases to support the test. The analysis was rerun with only Central (Ontario and Quebec) and Western (British Columbia, Alberta, Saskatchewan and Manitoba) regions being retained. A chi-square test for the goodness of fit revealed that both Central (37.8%) and Western (30.0%) regions were statistically comparable in terms of award rates ($\chi^2(1) = .001$, ns). Furthermore, mean rating scores for candidates applying from an Eastern ($M = 60.63$, $SD = 4.86$), Central ($M = 57.94$, $SD = 5.05$) or Western university ($M = 58.04$, $SD = 4.53$) illustrated the absence of any significant mean differences ($F(2, 270) = .875$, ns).

2.1.4 Institutional Funding

All institutions were classified as having either more than 15 million dollars in MRC funding or less than 15 million dollars in MRC funding. (Institutions classified as having more than 15 million dollars in MRC funding were as follows: Toronto, McGill,

No significant differences were found in award rates for:

- French language and English language candidates

- Applicants from Central or Western Canada   (the Eastern region had too few applicants for a statistical test)

- Applicants from institutions with more that $15 million in MRC funding versus those from institutions with less than $15 million

Montreal, British Columbia and Alberta.) Overall, there was no
significant difference in award rates between institutions with more
than 15 million dollars in MRC funding (34.7%) and those with less
than 15 million dollars (47.0%) in MRC funding ($\chi^2(1) = 3.66$, ns). On
average, scores for DRA candidates applying from an institution with
less than 15 million dollars in MRC funding ($\underline{M} = 58.88$, $\underline{SD} = 4.84$)
and more than 15 million dollars in MRC funding ($\underline{M} = 57.67$, $\underline{SD} =$
4.97) were comparable ($\underline{t}(271) = 1.867$, ns).

An examination of each specific selection criteria revealed
significant mean differences between candidates from institutions
classified as having either more (or less) than 15 million dollars in
MRC funding on measures of publication activity and graduate
performance.

- The mean score on **publication activity** for candidates coming
  from an institution with less than 15 million in MRC funding ($\underline{M}$
  $= 3.60$, $\underline{SD} = 1.36$) was higher than that for candidates coming
  from an institution with more than 15 million ($\underline{M} = 3.22$, $\underline{SD} =$
  1.38) in MRC funding ($\underline{t}(271) = 2.10$, $\underline{p} > .05$; $\underline{\omega}^2 = .012$).
  Institutional funding   accounted for 1.2% of the variability
  in publication activity.

- In comparison to candidates from institutions with more than 15
  million in MRC funding ($\underline{M} = 4.58$, $\underline{SD} = .81$), candidates
  applying from an institution with less than 15 million in MRC
  funding ($\underline{M} = 4.84$, $\underline{SD} = .78$) had higher mean scores on
  graduate academic performance ($\underline{t}(271) = 2.52$, $\underline{p} > .05$; $\underline{\omega}^2 =$
  .019). Institutional funding explained 1.9% of the variability in
  graduate academic performance.

Note that level of MRC funding in the applicant's institution
explained only a small percentage of the variability in the two
selection criteria. This finding suggested that significant differences
detected by t-tests may to some extent be an artefact of the large
sample size and may not indicate a meaningful difference.

2.1.5 Level of Graduate Study

A significant difference was detected between the award rates for candidates who were in a masters program (58.3%; n = 24) when applying for the DRA award and those who were in a doctoral program (36.5%; n = 249; $\chi^2(1)$ = 4.39, p < .05).   Furthermore, there was a significant difference in mean overall competition rating between candidates in a masters program (M = 60.63, SD = 4.91) and those in a doctoral program (M = 57.79, SD = 4.73) when applying ($t(271)$ = 2.71, p < .005; ω^2 = .023). Level of graduate study explained 2.3% of the variability in overall application scores.

> *Unexpectedly ...*
>
> The award rate for the 24 applicants who were still in a **masters program** was
>
> **higher**
>
> than the award rate for the 249 applicants who were already enrolled in **doctoral programs**.
>
> (58.3% vs 36.5%)

However, when differences were assessed by t-test on each specific selection criteria, candidates who were in a masters program (M = 6.72, SD = .56) when applying for the award only had significantly higher scores in comparison to those in a doctoral program (M = 6.34, SD = .90) on the measure of training program ($t(271)$ = 2.01, p < .05; ω^2 = .015). It appeared as though candidates applying from masters programs scored slightly (but not significantly) higher than candidates already in doctoral programs on the specific selection criteria which leadthe former group to obtain higher overall scores.

2.1.6 Research Orientation

10

A chi-square test for the goodness of fit indicated the difference in award rates between health research candidates (25.0%; n = 32) and biomedical research candidates (40.2%; n = 241) was not statistically significant ($\chi^2(1)$ = 2.775, ns). However, health research candidates had significantly lower mean scores (M = 55.78, SD = 5.38) compared to biomedical research candidates (M = 58.34, SD = 4.83), ($t(271)$ = 2.78, p < .01; ω^2 = .024). Orientation of the candidate's research accounted for 2.4% of the variability in overall rating. An examination of specific selection criteria indicated that:

- The mean score on **research resources** of the proposed doctoral supervisor for health research candidates (M = 2.87, SD = .82) was significantly lower than the mean score for biomedical research candidates (M = 3.48, SD = .57), ($t(271)$ = 5.35, p < .001; ω^2 = .092). Research orientation of the candidate explained 9.2% of the variability in research resources.

- Applications from health research candidates received lower scores (M = 3.10, SD = .63) on the **scientific activity** of the proposed doctoral supervisor than did applications from biomedical research candidates (M = 3.42, SD = .58), ($t(271)$ = 2.89, p < .005; ω^2 = .030). Research orientation of the candidate accounted for 3.0% of the variability in scientific activity.

- Also, applications from health research candidates (M = 2.67, SD = 1.0) scored significantly lower on **training record** of the proposed doctoral supervisor than did the applications of biomedical (M = 3.16, SD = .75) research candidates ($t(271)$ =

Differences Related to Research Orientation of the Candidate

Compared to applications from candidates in the biomedical area, those from candidates planning study in the social aspects of health had statistically significant lower scores overall and for all criteria relating to the doctoral training environment:

- **research resources of the doctoral supervisor**
- **scientific activity of the supervisor**
- **training record of the supervisor**
- **doctoral research project**

The difference seemed particularly strong with respect to the "research resources" variable. Research orientation of the candidate accounted for 9% of the variation in ratings of research resources.

3.32, p < .001; ω^2 = .035). Research orientation explained 3.5% of the variability in training record.

- Finally, health research candidates (M = 5.88, SD = 1.2) scored significantly lower on the proposed **training program** for the candidate in comparison to biomedical (M = 6.43, SD = .82) research candidates (t(271) = 3.38, p < .001; ω^2 = .037). Research orientation of the candidate accounted for 3.7% of the variability in training program.

It is worth noting that when scores for all four criteria are considered as a block
(e.g., research training environment), there was a statistically significant difference in mean scores between health (M = 14.52, SD = 2.97) research candidates and biomedical (M = 16.49, SD = 2.13) research candidates (t(271) = 4.66, p < .001; ω^2 = .070). In summary, health research candidates scored comparably with biomedical research candidates on measures of characteristics and abilities, initial research activity and academic performance with but not as well on measures of research training environment.

2. 2   Examination of Fairness Focused on Sponsors

2.2.l   Gender

Each candidate was supported by three sponsors.   In 94.1% of the applications, at least one of the sponsors was also the student's doctoral research supervisor. Unfortunately, the gender of other sponsors (those who were not doctoral supervisors) was not known. For this reason, the analysis by gender focused on the supervisors.

With respect to the applicants' sponsors:

No differences between scoring of applicants by female or male sponsors

but ...

significant mean differences between the scores provided by doctoral research supervisors and other sponsors.

(26.3 by doctoral supervisors vs 25.7 by other sponsors)

To assess whether male and female supervisors differed in their scoring of (a) applicants generally or (b) male and female DRA candidates specifically, a 2X2 between-groups analysis of variance was performed. (Note that for those applications where the supervisor had not included a sponsor's report, the case was not included). Overall, it appeared that there was no significant difference in the overall scoring of sponsor-assessed selection criteria by male ($\underline{M}$ = 25.89, $\underline{SD}$ = 3.32) and female ($\underline{M}$ = 25.99, $\underline{SD}$ = 2.41) supervisors ($\underline{F}$(1, 254) = .459, ns; $\underline{\eta}^2$ = .002).

Furthermore, the interaction between the gender of the supervisor and the gender of the DRA candidate was not significant ($\underline{F}$(1, 254) = .016, ns; $\underline{\eta}^2$ = .001). That is, male and female supervisors do not score male and female candidates differently on the sponsor's report form. A series of t-tests indicated that male and female supervisors do not score specific sponsor-assessed selection criteria differently.

2.2.2 Supervisors' Versus Other Sponsors' Ratings

A series of t-tests were performed to see if the doctoral supervisor tended to score the candidate higher than the two other sponsors. (Only the cases where the supervisor provided a sponsor report were included ($\underline{n}$ = 256) in the analysis. Also, in six cases the DRA candidate had two supervisors who provided a sponsor report. In those cases, the average rating for the two supervisors was compared to the rating of the single sponsor.)

- For the overall score of criteria on the sponsors' report form, supervisors ($\underline{M}$ = 26.33, $\underline{SD}$ = 1.82) compared to other sponsors ($\underline{M}$ = 25.69, $\underline{SD}$ = 1.52) tended to score candidates higher ($\underline{t}$(510) = 4.22, $\underline{p}$ < .001; $\underline{\eta}^2$ = .032).

- Supervisors, in comparison to other sponsors, tended to score candidates higher on all of the individual selection criteria, except for independence of the candidate (See Table 5).

However, the effect sizes were relatively small. That is, the type of work-related relationship the sponsor had with the candidate (doctoral supervisor or not) did not explain a substantial proportion of variability in overall and selection criteria scores.

2.2.3 First Academic Appointment of Supervisor

A file review was conducted on a random sample of 50 DRA applications to extract information on the first academic appointment of the proposed doctoral supervisor. All doctoral supervisors were classified into one of the two categories: received first academic appointment prior to 1990, received first academic appointment in 1990 or later.   An examination of the selection criteria which assessed the scientific activity, research resources and training record of the supervisor indicated that:

> *"Younger supervisors"* were defined as those whose first academic appointment was in 1990 or later. They had **lower scores** on measures of:
>
> - **scientific activity** (3.1 vs. 3.6)
>
> - **research resources** (3.2 vs. 3.5)
> **but**
> there was no statistical significant difference between their scores and those of "older" supervisors on a measure of:
>
> - training record

- Supervisors who received their first academic appointment in 1990 or later had significantly lower scores on **scientific activity** (M = 3.1, SD = .59) than supervisors who received their appointment prior to 1990 (M = 3.6, SD = .45), (t(48) = 2.91, p < .005; ω^2 = .13). First academic appointment accounted for 13% of the variability in scientific activity.

- Supervisors (M = 3.2, SD = .56) who received their first academic appointment in 1990 or later had significantly lower scores on **research resources** than supervisors (M = 3.5, SD = .44) who received their appointment prior to 1990 (t(48) =

2.91, p < .005; $\underline{\omega}^2$ = .07). First academic appointment accounted for 7% of the variability in research resources.

• Supervisors ($\underline{M}$ = 2.9, $\underline{SD}$ = .54) who received their first academic appointment in 1990 or later scored comparably on **training record** with supervisors ($\underline{M}$ = 3.2, $\underline{SD}$ = .85) who received their appointment prior to 1990 ($\underline{t}$(48) = 1.38, ns).

2.3   Examination of Fairness Focused on Committee Members

2.3.1.   Gender

To assess whether male and female DRA committee members differed in their scoring of (a) applicants generally or (b) male and female candidates specifically, a 2X2 between-groups analysis of variance was performed. It appeared as though female committee members tended to give higher scores ($\underline{M}$ = 60.06, $\underline{SD}$ = 5.26) in comparison to male committee members ($\underline{M}$ = 58.18, $\underline{SD}$ = 5.44) on the DRA application ($\underline{F}$(1, 542) = 7.65, $\underline{p}$ < .01; $\underline{\eta}^2$ = .028). Gender of committee members accounted for 2.8% of the variability in the DRA application. However, the interaction between gender of the committee member and gender of the applicant was not significant ($\underline{F}$(1, 542) = 1.30, ns). That is, male and female DRA committee members scored male and female candidates in a similar fashion.

Given that male and female DRA committee members tended to score the application differently, a series of t-tests were performed to identify those criteria on which they differed. The statistical tests suggested that out of the eight selection criteria assessed only by committee members (i.e., not including sponsor-assessed criteria), differences related to committee member gender were detected for three: candidate`s publication activity, other research activity of the candidate and training program for the candidate.

• In comparison to male committee members' rating of candidate's **publication activity** ($\underline{M}$ = 3.17, $\underline{SD}$ = 1.53), female

committee members tended to give higher scores ($\underline{M}$ = 3.61, $\underline{SD}$ = 1.46), ($\underline{t}$(544) = 2.40, $\underline{p}$ < .05; $\underline{\omega}^2$ = .020). Gender accounted for 2.0% of the variability in publication activity.

- Furthermore, female committee members ($\underline{M}$ = 3.96, $\underline{SD}$ = 1.32) appeared to give higher scores on candidate`s **other research activity** in comparison to male ($\underline{M}$ = 3.68, $\underline{SD}$ = 1.32) committee members ($\underline{t}$(544) = 2.40, $\underline{p}$ < .005; $\underline{\omega}^2$ = .009). Gender explained 0.9% of the variability in other research activity.

- Lastly, in comparison to male committee members scores for the candidate's **training program** ($\underline{M}$ = 6.26, $\underline{SD}$ = 1.086), female committee members also tended to give higher scores ($\underline{M}$ = 6.55, $\underline{SD}$ = 1.194), ($\underline{t}$(544) = 2.91, $\underline{p}$ < .005; $\underline{\omega}^2$ = .01) but this difference accounted for only 1.0 % of the variability in training program.

Although t-tests revealed statistically significant differences between female and male committee members on means of overall total, publication activity, other research activity and training program, an examination of the effect sizes revealed that the differences between male and female committee member mean scores were only slight. Mean differences related to committee member's gender may have been found significant because of the large sample size.

There were some statistically significant differences between the mean scores assigned by female and male committee members:

	Female committee members	Male committee members
Overall score	Higher 60.1	58.2
Publication activity	Higher 3.6	3.2
Other research activity	Higher 4.0	3.7
Training program	Higher 6.6	6.3

but ...

There was no relationship between gender of reviewers and their ratings of male and female applicants

and

the effect of committee member gender on the variability of scores was small.

16

For further analysis, DRA candidates were placed in three groups according to the gender of the two committee members that reviewed each candidate. The groups are as follows: female-female reviewer (FF: n = 31), female-male reviewers (FM: n = 100) and male-male reviewers (MM: n = 142). Overall, there were no statistically significant differences in award rates for candidates reviewed by FF (41.9%), FM (38.7%) and MM (37.0%) committee member gender pairs ($\chi^2(2)$ = .253, ns).

To assess whether the three committee member gender pairs differed in their scoring of (a) applicants generally or (b) male and female candidates, a 3X2 between-groups analysis of variance was performed. Overall, there were no significant differences in the overall mean scores of applicants reviewed by FF ($\underline{M}$ = 58.82, $\underline{SD}$ = 5.68) committee member pairs, FM ($\underline{M}$ = 58.32, $\underline{SD}$ = 4.50) committee member pairs and MM ($\underline{M}$ = 57.39, $\underline{SD}$ = 5.30) committee member pairs ($\underline{F}(2, 267)$ = .859, ns; $\underline{\eta}^2$ = .009). Moreover, the interaction between gender of the candidate and gender composition of the committee member pair was not significant ($\underline{F}(2,267)$ = .447, ns; $\underline{\eta}^2$ = .003). That is, the committee member gender pairs tended not to score male and female candidates differently.

2.3.2. Language Preference

To assess whether French language preference and English language preference DRA committee members differed in their scoring of (a) applicants generally and (b) French and English applications, a 2X2 between-groups analysis of variance was conducted. Overall, no significant

> **No significant mean differences** were found between ratings provided by committee members who indicated preference for French language correspondence and committee members who indicated a preference for English language correspondence.

difference existed between the overall rating produced by French language (M = 57.58, SD = 5.17) and English language (M = 58.11, SD = 5.86) committee members (F(1, 542) = .376, ns). Likewise, the interaction between language preference of the committee member and language of the DRA application was not significant (F(1, 542) = .188, ns). That is, French language and English language DRA committee members tended to score French and English applications the same way.

2.3.3 Research Orientation

To assess whether DRA committee members, who were either health or biomedical researchers, differed in their scoring of (a) applicants generally or (b) health and biomedical candidates, a 2X2 between-groups analysis of variance was performed. It appeared as though committee members who were in health research (M = 59.12, SD = 5.39) tended to give higher scores in comparison to committee members who were in biomedical research(M = 57.73, SD = 5.66) on the DRA application (F(1, 542) = 8.10, p < .005; n^2 = .015).   Although, health and biomedical researchers rated candidates who were in either health or biomedical research in a similar manner (F(1, 542) = 0.12, ns). For example, when health research candidates were isolated (n = 32), no significant mean differences were detected for overall score and individual selection criteria between health and biomedical research committee members (see Table 6).

> On average, health research committee members tended to score candidates higher than biomedical research committee members
>
> **but**
>
> health and biomedical research committee members scored health and biomedical research candidates in a comparable manner.

To identify the selection criteria which health researchers tended to score higher than biomedical research committee members, a series of t-test were conducted. Significant group differences were

18

detected for three: training record, undergraduate academic performance and graduate performance. That is:

- In comparison to biomedical research committee members' rating of the supervisor's **training record** (M = 3.05, SD = .93), health research committee members (M = 3.28, SD = .87) tended to give higher scores (t(5) = 2.43, p < .05; ω^2 = .01). Research orientation of committee member accounted for 1.0% of the variability in training record.

- Furthermore, health research committee members (M = 4.55, SD = 1.04) appeared to give higher scores on candidate`s **undergraduate academic performance** in comparison to biomedical research committee members (M = 4.29, SD = 1.05) committee members (t(544) = 2.42, p < .05; ω^2 = .01). Gender explained 1.0% of the variability in undergraduate academic performance.

- Lastly, in comparison to biomedical research committee members scores for the candidate's **graduate academic performance** (M = 4.58, SD = .96), health research committee members (M = 4.95, SD = .86) also tended to give higher scores (t(544) = 3.90, p < .001; ω^2 = .025) but this difference accounted for only   2.5 % of the variability in graduate academic performance.

19

3.   INTER-RATER RELIABILITY

3.1   Intraclass Correlations

The intraclass correlation coefficient is a measure of agreement between raters. It ranges from 0 to 1, where 1 represents absolute agreement. The inter-rater reliability of the DRA rating form was .72 with a mean difference of 5.63 ($\underline{SD}$ = 4.27) between committee members. For the individual selection criteria (See Table 7), inter-rater reliability was:

- good for candidate's **publication activity** (.80), supervisor's **training record** (.68) and candidate's **undergraduate record** (.64)

- fair for **scientific activity** of doctoral environment (.55), **research resources** of doctoral laboratory (.55), **other research activity** of candidate (.54) and candidate's **graduate performance** (.53)

- poor for candidate's **training program** (.36) and the **adjustment score** by which committee members could reflect their views on sponsors' ratings (.27)

3.2   Committee Members' Use of the Rating Scale

All DRA committee members were ranked, in descending order, in terms of the percentage of applications for which they gave a higher score than their companion reviewer.

Committee members were ranked according to the percentage of applications for which they were the highest reviewer.

The ranked list, with committee members identified only by a code number, is provided in Appendix D.

- Reviewers in the top 30% of the list were categorized as high raters (H); and

- those in the bottom 30%, low raters (L).

The mean scores for applications reviewed by two high markers was 60.1 and by two low markers was 56.4 but the difference was not statistically significant.

Committee members who were in the top 30% of the list were categorized as "high raters" and those who were in the bottom 30% of the list were categorized as "low raters". The following analysis focused on DRA candidates who had their application reviewed by either two high raters (HH) or two low raters (LL). Overall, there was no statistically significant difference in the mean overall scores of the HH ($\underline{M}$ = 60.07, $\underline{SD}$ = 5.56) and the LL ($\underline{M}$ = 56.36, $\underline{SD}$ = 5.93) committee member pairs ($\underline{t}$(37) = 2.02, $\underline{ns}$). Furthermore, an examination of DRA award rates for DRA candidates who were rated by either HH or LL pairs of committee members revealed no significant differences in award rates ($\chi^2(1)$ = 3.143, ns).

4.   SELECTION CRITERIA

4.1   *Internal Consistency*

According to Cronbach's standardized coefficient alpha, the DRA rating form demonstrated an adequate level of internal consistency (α = .75). The same was true for the sponsor-assessed criteria (α = .79) and committee member-assessed criteria (α = .71). An examination of the Cronbach's alphas that would result if selection criteria were deleted indicated that no significant improvements could be made to the reliabilities by the elimination of items.

4.2   *Relative Importance of the Selection Criteria*

It might be expected that a DRA candidate who had a high overall score should also have had high scores on the individual selection criteria. However, it is also plausible that DRA recipients and non-recipients might tend to have similar scores on some of the selection criteria.

> *A regression analysis (with scores for individual criteria as independent variables) indicated that:*
>
> - personal characteristics of the candidate
> - academic performance
>
> are somewhat less important than other variables in determining total score.

21

One might view these criteria as less important in the context of this competition. T-tests revealed that the awardee group had statistically significant higher scores than the non-awardee group for all the individual selection criteria.

Another way of considering the relative importance of the DRA selection criteria, is to compare their weights in a regression equation where the dependent variable is DRA total score. As expected, the statistic associated with such a regression equation suggests that overall score can be significantly predicted on the combined basis of the DRA selection criteria ($\underline{F}$ (15, 257) = 1278.58, $\underline{p}$ < .001). Zero-order correlations between the reviewer-assessed criteria and total score indicated:

- a **strong positive relationship** between total score and candidates' publication activity and other research activity, scientific activity and research resources of the doctoral training environment, supervisor's training record and candidates' training program.

- a **moderate positive relationship** between total score and undergraduate grades, graduate performance and all sponsor-assessed selection criteria

The standardized regression coefficients (β) also indicate the importance of criteria and, unlike the zero-order correlations, take into account variation in the scaling of criteria. An examination of the standardized betas revealed that some selection criteria weighted more heavily in the prediction of total score than others. For example, candidate's other research activity ($\underline{\beta}$ = .251, $\underline{p}$ <.001), publication activity ($\underline{\beta}$ = .276, $\underline{p}$ < .001) and training program ($\underline{\beta}$ = .198, $\underline{p}$ < .001) weighted into the regression equation more heavily than independence ($\underline{\beta}$ = .052, $\underline{p}$ < .001), interest in discovery ($\underline{\beta}$ = .050, $\underline{p}$ < .001) and critical thinking ($\underline{\beta}$ = .047, $\underline{p}$ < .001). Generally, the standardized regression coefficients were higher for the reviewer-assessed criteria than the sponsor-assessed criteria.

22

4.3 Natural Grouping of Selection Criteria

Using their intuition and experience, designers of the DRA assessment form had grouped criteria into four categories: personal characteristics of the candidate, initial research activity of the candidate, research training environment and academic performance of the candidate. To examine whether or not the rating data supported such a classification, or pointed to some other grouping of criteria, a principal components analysis (PCA) with a varimax rotation was conducted.

> *A statistical analysis of the covariances among the criteria indicated the presence of three factors (or conceptual groupings) among them:*
>
> The three factors are:
>
> - characteristics and abilities
>
> - research training environment
>
> - candidate's achievements

The analysis suggested that three factors underlie the 15 selection criteria. These factors appeared to be: characteristics and abilities of the candidate (Factor 1), research training environment (Factor 2), and candidate's achievements (Factor 3). Overall, the factor solution explained 48.2% of the variability (i.e., spread in scores).

- Factor 1, which seemed to capture **"characteristics and abilities of the candidate"**, illustrated the association between a candidate`s critical thinking, independence, perseverance, originality, organizational skills, interest in discovery and research ability.

- Factor 2, **"research training environment"**, was defined by the following core variables: scientific activity, research resources, training record of the doctoral supervisor and training program for the candidate.

- Factor 3, which appeared to tap into the **"candidate's achievements",** demonstrated the relationship between

23

publication activity, other research activity, undergraduate academic performance and academic performance as a graduate student.

Summary Table of Findings

Area of Assessment	Gender of candidate (A)	Language preference of candidate (A)	Region (A)	Institutional funding (A)	Level of graduate study (A)	Research orientation of the candidate (A)	Gender of the advisor (A)	Role of supervisor (A)	Supervisor academic appointment (A)
Characteristics	-	-	-	-	-	-	-	Supervisor ↑	NA
Critical thinking	-	-	-	-	-	-	-	Supervisor ↑	NA
Independence	-	-	-	-	-	-	-	Supervisor ↑	NA
Perseverance	Females ↑	-	-	-	-	-	-	Supervisor ↑	NA
Originality	Males ↑	-	-	-	-	-	-	Supervisor ↑	NA
Organizational skills	Females ↑	-	-	-	-	-	-	Supervisor ↑	NA
Interest in discovery	Males ↑	-	-	-	-	-	-	Supervisor ↑	NA
Research ability	-	-	-	-	M ↑	-	-	Supervisor ↑	NA
Adjustment score	-	-	-	-	-	-	-	NA	NA
Initial Research Activity	-	-	-	--	-	-	-	NA	NA
Publication activity	-	-	-	less than 15 ↑	-	-	NA	NA	NA
Other research activity	-	-	-	-	-	-	NA	NA	NA
Research Training Environment	-	-	-	-	-	BIO ↑	NA	NA	NA
Scientific activity	-	-	-	-	-	BIO ↑	NA	NA	PRIOR
Research resources	-	-	-	-	-	BIO ↑	NA	NA	PRIOR
Training record	-	-	-	-	-	BIO ↑	NA	NA	-
Training program	-	-	-	-	M ↑	BIO ↑	NA	NA	NA
Academic Per.	-	-	-	-	-	-	NA	NA	NA
Undergrad performance	-	-	-	-	-	-	NA	NA	NA
Graduate performance	-	-	-	less than 15 ↑	-	-	NA	NA	NA
Total	-	-	-	-	-	BIO	NA	NA	-

- **Groups:**
  - *Gender :* Female, Male
  - *Language preference:* French, English
  - *Research orientation:* Biomedical researcher (BIO), Health researcher
  - *Region:* Eastern, Central, Western
  - *Institutional funding:* More than 15 million dollars in MRC funding, Less than 15 million dollars in MRC funding
  - *Level of graduate study :* M (Master), D (Doctorate)
  - *Role of sponsor :* Supervisor, Not supervisor
  - *First academic appointment of the supervisor:* Prior to 1990, in 1990 or later

- A: looking at mean group differences
- (group) ↑: (group) had the higher mean score
- NA: Not Applicable
- Role of sponsor: characteristics and abilities total does not include the adjustment score provided by the committee members.
- A full description of the results is located in the technical report.

Area of Assessment	Gender of committee member (A)	Gender of committee member pair (A)	Language preference of committee member (A)	Research orientation of committee member (A)	Inter-rater reliability	Internal consistency	Importance in predicting overall score	Did recipients score higher than non-recipients? (A)
Characteristics	-	-	-	-	GOOD	GOOD	NA	YES
Critical thinking	NA	NA	-	NA	NA	NA	MODERATE	YES
Independence	NA	NA	-	NA	NA	NA	MODERATE	YES
Perseverance	NA	NA	-	NA	NA	NA	MODERATE	YES
Originality	NA	NA	--	NA	NA	NA	MODERATE	YES
Organizational skills	NA	NA	-	NA	NA	NA	MODERATE	YES
Interest in discovery	NA	NA	-	NA	NA	NA	MODERATE	YES
Research ability	NA	NA	-	NA	NA	NA	MODERATE	YES
Adjustment score	-	-	-	NA	POOR	NA	NA	NA
Initial Research Activity	-	-	-	-	GOOD	GOOD	NA	YES
Publication activity	Females ↑	-	-	-	GOOD	NA	HIGH	YES
Other research activity	Females ↑	-	-	-	FAIR	NA	HIGH	YES
Research Training Environment		-	-	-	GOOD	GOOD	NA	YES
Scientific activity		-	-	-	FAIR	NA	HIGH	YES
Research resources		-	-	-	FAIR	NA	HIGH	YES
Training record		-	-	-	GOOD	NA	HIGH	YES
Training program	Females ↑	-	-	-	GOOD	NA	HIGH	YES
Academic Per.		-	-	-	GOOD	POOR	NA	YES
Undergrad performance		-	-	-	GOOD	NA	HIGH	YES
Graduate performance		-	-	-	FAIR	NA	HIGH	YES
Total	Females ↑	-	-	--	GOOD	GOOD	HIGH	YES

Part B

Survey of

DRA Committee Members

INTRODUCTION

A survey of DRA committee members was designed to elicit committee members' perceptions of the selection process and criteria. The questionnaire used for the survey included a series of closed-ended questions with a free-response section at the end of each question.

Distribution and Methodology

Prior to the administration of the survey, the questionnaire was pre-tested by members of the DRA core committee. The questionnaire was revised, then in July 1998, was circulated to all 34 DRA committee members by fax. By mid July, 19 completed surveys had been returned to the MRC. A follow-up request resulted in a further 10 returns. The overall response rate was thus 85.3% (29/34).

Quantitative data was coded and entered into an SPSS file as received. Committee members' comments were transcribed and a content analysis conducted at the beginning of August. Every effort was made to extract the essence of the information without editorializing.

> **Terminology note:** we will use the terms "committee member" and "reviewer" interchangeably.

RESULTS

2.1   Experience with the Review Process

Question 1*:* ***On average, how long did it take to review each application?***

On average, it took committee members 40 minutes ($\underline{SD}$ = 24) to review and evaluate a DRA application.   The length of review time ranged from 10 to 120   minutes. The most frequent response was 30 minutes (36%; 10/28).

Question 2:   *Did you have any difficulty in meeting the February 6, 1998 DRA application review deadline?*

Most DRA committee members (93%; 27/29) had no difficulty in meeting the application review deadline.

Question 3:   *Did you find it necessary to consult colleagues in other research areas concerning research proposals, publication records or other variables influenced by the field of science involved?*

Most committee members (79%; 23/29) did not find it necessary to consult colleagues. However, six committee members did with the number of consults ranging from 2 to 5. Four committee members provided a comment. Sample quotes:

> *"I did have problems with some applications. For these I did a bit of on-line research."*

> *"About 5 applications were not basic science and I was not aware of   the merit/achievement milestones that would be normal for those candidates."*

Question 4:   *Committee members were asked to use only whole numbers in   their ratings. Should members have the option of using intermediate scores with one decimal place?*

The following are the committee members responses:

Response Category:	n =29	%
• yes, but only .5	18	62.1
• yes, .1 to .9	7	24.1
• no	4	13.8

It appears as though committee members (86%; 25/29) are in favour of the   intermediate score option with the majority endorsing the .5 over the .1 to .9 option[1]. The following is representative sample of comments:

> *"If necessary yes, but only .5. In fact my decision process was made   easier because I could only use whole numbers".*

> *"I found the whole number scale to be too crude for discriminating   between closely matched applications".*

Question 5: ***Did the clarity of writing in the description of the candidate's Training project influence your assessment of the candidate?***

Most committee members (83%; 24/29) indicated that the clarity of write-up for the candidate's training project had an impact on their assessment. The following themes emerged from a review of their comments (n = 21):

- the writing provided an indication of clarity of thinking, preparation skills and/or communication skills (5).

- clarity of writing became an important issue when the project was not in the area of the committee members' expertise (4).

- a clearly written project is easier to assess (3).

---

[1]   In fact, while use of whole numbers was requested, a few committee members used intermediate scores in the first competition.

- reviewer concern that the training project was written by the supervisor (2).

- reviewer was slightly lenient if application was not written in the candidate's first language (2).

- clarity of writing indicates the candidate/supervisor interaction (1).

- other (4).

A representative sample of comments from committee members:

"I believe that good communication is part of the research. Candidates of their level should be able to concisely describe what it is they want to do, why they want to do it and how they intend to do it."

"Some of the projects were not in my area of expertise. Therefore, the clarity was important for me."

"[In] many cases the training project was obviously written by the supervisor (a hopefully more experienced writer and thus perhaps unfairly overrated)."

"All applications that I examined were of adequate clarity."

Question 6: ***If you reviewed any applications that were written in your second language, were you at all concerned about your reading comprehension?***

The majority of committee members (62%; 18/29) only reviewed and evaluated applications written in their first language. Of the 10 committee members who indicated that they reviewed applications in their second language, 7 were not concerned but 2 were somewhat concerned about their reading comprehension. One committee member indicated a major concern and provided the following comment:

"I'm not fluent in French and I had the application translated by a non-scientist."

Question 7:   **Do you think that the guidelines provided by the MRC are sufficiently explicit to cover all conflict of interest situations that a reviewer might encounter?**

All reviewers (29/29) thought that the guidelines were sufficiently explicit to cover all potential conflict of interest situations. No comments were made.

2.2   Satisfaction with the Criteria and Weights

Question 8:   **Do you feel that the some of the selection criteria should be eliminated?**

A majority of committee members (66%; 19/28) felt that all selection criteria currently employed in the assessment of candidates should remain. The remaining 9 committee members commented on their concerns with specific criteria.

- sponsor's reports (2 commented, sample comment provided below).

  "All [sponsors] said that the candidate was the best that they had ever seen and were seldom critical. They should be used as a guide but not used in the rating."

- other research activity (2 commented, sample comment provided below).

  "*Is a vague category because candidates do not seem to have much other research activity and spot on the application which allows them to elaborate on this."*

- training record of the supervisor (1 commented).

  "Eliminate, this is too hard to evaluate because new investigators will always have fewer or no students (and quality of the environment program [is] included in other categories). In essence, one can only mark down and this is too hard to evaluate. There are too many extenuating circumstances (e.g., new investigators, small graduate programs etc.)."

- research ability (1 commented).

  "This is highly correlated with interest in discovery and is redundant."

- perseverance (1 commented).

  "Difficult to assess and [of] questionable relevance."

Question 9: **Are there any additional selection criteria that you think should be considered?**

Eight committee members (29%; 8/28) thought additional criteria should be included in the assessment of DRA candidates. The recommended selection criteria are:

- scientific writing/presentation skills (2).

  "These are essential tools for a research scientist and such skills vary tremendously between students, or at least in my experience."

- level of interest in research and future career plans (1).

  "The candidate's level of interest in research and future plans for their careers. The emphasis should be put on the students wanting to continue in research."

- reviewer's overall impression(1).

  "How about an overall impression grade. I sometimes got the feeling a student was more or less deserving but could not put my finger on it."

- breadth and depth of knowledge (1).

  "Sponsors could be asked to evaluate the candidate's breadth and depth of knowledge in the field of study."

- other (3).

  "Under training program, opportunity to evaluate the teaching skills should be given."

> *"My concern is the placing of social science and physical science candidates together. They must be considered separately."*

> *"New students may be assessed separately from those who are in second or third year of the Ph.D program."*

General comments (not related to specific criteria):

> *"I think that the process we used is still "marble" counting, rather than a global assessment of the application and his or her training program."*

> *" I felt that it was a comprehensive list of criteria."*

Question 10:   ***Please indicate your views on the appropriateness of the weight that each selection criteria carries in the final score. If you recommend a different weight, please indicate the recommended value.***

Selection Criteria[2]	Committee Members' Views on Current Weight				Direction of Weighting Change and Number of Times Suggested[3]	
	# of responses	Weight is appropriate	Undecided	Weight is inappropriate	Increase the weight	Decrease the weight
Critical thinking	27/29	19 (70%)	0 (0%)	8 (30%)	3	5
Independence	27/29	19 (70%)	0 (0%)	8 (30%)	2	6
Perseverance	27/29	21 (78%)	0 (0%)	6 (22%)	0	5
Originality	27/29	20 (74%)	2 (7%)	5 (19%)	1	4
Organizational skills	27/29	21 (78%)	0 (0%)	6 (22%)	0	5
Interest in discovery	27/29	20 (74%)	1 (4%)	6 (22%)	1	4

---

[2]   C = candidate, S = proposed doctoral supervisor

[3]   Not all committee members who felt a weight was inappropriate provided a suggestion. Moreover, committee members who were undecided about a weight occasionally recommended a new weighting.

Research ability	27/29	18 (67%)	1 (4%)	8 (30%)	2	6
Adjustment score	26/29	20 (77%)	0 (0%)	6 (23%)	1	3
Publication act. [c]	27/29	16 (59%)	2 (7%)	9 (33%)	4	1
Other research activity [c]	27/29	18 (67%)	7 (26%)	2 (7%)	1	1
Scientific activity [s]	26/29	21 (81%)	1 (4%)	4 (15%)	0	3
Rsrch resources [s]	27/29	20 (74%)	0 (0%)	7 (26%)	0	5
Training record [s]	27/29	20 (74%)	1 (4%)	6 (22%)	0	4
Training prog. [c]	26/29	20 (77%)	1 (4%)	5 (19%)	2	3
Undergraduate performance [c]	27/29	17 (63%)	8 (30%)	2 (7%)	3	3
Graduate performance [c]	27/29	15 (56%)	2 (7%)	10 (37%)	6	2

The following is a representative sample of comments supplied by committee members:

> *"Reports from the sponsors were always high. Weakness of the students were not mentioned in 99% of the sponsor letters. I recommend reduction of weights."*

> *"Publication activity should weight more than other research activity."*

> *"Publications could be stressed more but with the provision that the candidates can explain why they haven't been able to publish if they are beyond 1-2 years of grad study."*

> *"Other research activity. This category is very vague and poorly described and yet it contains similar weighting as the publication category. I would prefer a lower weighting."*

> *"I didn't like strong supervisors carrying the weight of poor students."*

> *"Research resources. I think that this should be a yes/no category. There is no evidence that students receive a better education if their supervisor has more money (i.e., lots of grant funds vs small grants). As long as the reviewer can assess that the supervisor has enough funds to support the project, the mark should be yes (1). If not, then no (0)."*

"Undergraduate academic record should weight more than graduate academic performance or graduate performance could be eliminated since students usually have only a few graduate courses (if any) and graduate course transcripts vary (one could just have one category to include course work)."

"Graduate academic performance is usually, if not always, quite high in basic sciences."

Question 11: **During your review of applications, did you find that any of the selection criteria were particularly difficult to assess?**

Many committee members (75%; 21/28) found at least one selection criteria difficult to assess. The following is a list of criteria where at least one committee member noted assessment difficulty[4]:

- sponsor's reports (6).

  "I did not find the letters from referees to be very critical for the purpose of evaluating the student."

- candidate's publication activity (5).

  "There is a variety of types of publications and awards. Should they weight equally? Is one 4th author published paper worth as much as a 1st author paper in preparation? etc."

- other research activity of the candidate (5).

  "What should we classify as other research activity? Should it be seminars, workshops, journal clubs, abstracts; all any or none?"

- training program for the candidate (4).

  "Some supervisor's filled this in, some did not take it seriously, some let the students do it."

---

[4]   Some committee members commented on more than one criteria.

- undergraduate academic performance (4).

     "Very difficult when assessing foreign transcripts."

- graduate academic performance (3).

     "Had very little information."

- training environment (1).

     "In some cases the training environment and its suitability was not entirely clear to me."

2.3   Satisfaction with the Rating Guide

Question 12: ***Were there any particular criteria for which the rating guidelines need to be improved?***

Eleven committee members (41%; 11/27) felt the rating guidelines should be improved. Seven committee members provided comments specifically addressing this question. Other comments which referred to changes in the weighting of criteria, were merged with the comments on 10 (appropriateness of weights). The seven comments that clearly called for improvement to the rating guide focused on the following criteria:

- academic performance as a graduate student (1).

- sponsors' reports (1).

- research training environment (1).

- other, not relating to specific criteria (2):

1.   sort by social sciences/biological sciences and new graduate/advanced graduate (1).

   2.   give examples (1).

A sample of quotes provided by committee members:

> "*Grading the supervisors training record by counting former trainees gives an advantage to applicants with older supervisors. The guide should clearly Indicate that the supervisors years of tenure should be considered when evaluating this category.*"

> "*Reports from sponsors were very subjective and were not critical. They should be presented to the reviewer as a guide and a means of clarification, but they should not be rated since all the scores were very similar.*

> "*I thought that the instructions were good.*"

2.4 Reviewer Differences

Question 13: **To be fair to applicants, the MRC sought a third opinion when the discrepancy between the two committee members' scores was large. After all the members' scores had been received, the MRC identified those cases in which the difference between the two reviewers was greater than the mean difference for all applications (4.5) plus one standard deviation from that mean (3.5). If narrowing the gap could have lead to a significant change in the status of an application, a third opinion was sought.**

You are most welcome to suggest a better method (or methods) for dealing with major discrepancies in scores.

Eight of the 24 committee members provided a response to the issue raised in the question. All responses fell into one of the following categories:

- the method for dealing with discrepancies is appropriate (3).

- there needs to be more communication between committee members (3).

- the method is appropriate but should be complemented by more communication between committee members (2).

Suggested forms of communication for dealing with discrepancies included:

- reviewers consult each other to reevaluate a discrepant score (3).

- a mediator (e.g., MRC) can call committee members and discuss particular aspects of why a score was given (2).

The following is a representative sample of committee member comments:

"Perhaps a dialogue between the two committee members involved might lead to a consensus rating - this would avoid a third review."

"If on the other hand there is a continuing discrepancy between the scores as in a mid range value is obtained there is obviously something different between them and maybe a communication between individuals may result in a greater consensus in the opinion. This actually could be done by a mediator calling the assessors and discussing the particular aspects of why a score was given and this may come up with a more considered opinion."

"It would have been useful to me to know whether I personally was involved in such cases."

Question 14: ***Using the competition database, we will be able to calculate the proportion of a reviewer's set of applications for which she/he was the higher reviewer. Would you be interested in knowing the proportion of applications for which your score was the highest?***

A vast majority of committee members (90%; 26/29) were interested in knowing the proportion of applications for which they were the highest reviewer.

Question 15: ***In general, how do you perceive yourself as a reviewer?***

Half of the group of committee members view themselves as neither critical nor lenient. Committee members' responses are as follows:

Response Category:	*n = 26*	*%*
very lenient	0	0.0
lenient	3	11.5
somewhat lenient*	3	11.5
average	13	50.0%
somewhat critical*	1	3.8
critical	6	23.1
very critical	0	0.0

    *   these categories were created for committee members who
       circled both "lenient" and "critical" or both " critical and average".

A representational sample of committee members' comments are the following:

> *"Who knows? I think that I am lenient, I am told I am not."*

> *"Most of the applications were outside of my area of expertise, so I could not be as critical as I would like to be."*

> *"I have no idea. That's why I would be interested in the results from above."*

2.5   Interaction and Communication

Question 16: ***In your review, what are the three main advantages and/or disadvantages of a virtual committee approach (i.e., no meeting of the full committee) as is used for the DRA review process.***

1.   Advantages of the DRA selection process

Committee members appear to perceive two main advantages of the DRA selection process: economy of time, and cost effectiveness. Responses were grouped within the following themes:

- economy of reviewer's time (19).

- cost effectiveness (18).

- independence of review (5).

- not having to travel (4).

- more objective (4).

- convenience (2).

2.   Disadvantages of the DRA selection process

The following is a categorization of committee member responses:

- no feedback/discussion (10).

- discrepancies between committee members (8).

- no opportunity to meet colleagues (5) / impersonal (2).

- no opportunity to modify score (5).

- impersonal (2).

- ineffective use of committee member expertise (2).

- other (2).

Lack of discussion/feedback and the discrepancies between ratings of committee members appear to be the two main disadvantages of the DRA selection process as perceived by committee members.

Question 17*: **In your opinion, would it be useful to have a forum to discuss evaluation issues with other DRA committee members?***

In response to this question, 41% (12/29) of committee members indicated yes, 28% (8/29) indicated no and 32% (9/29) were undecided. Four types of desired forum emerged from comments supplied by 17 committee members:

1. conference calls (9):

   • to resolve discrepancies between committee members (4).

   • for a general discussion among committee members (3).

   • for ranking of very good students (1).

   • for core committee member discussions of issues (1).

2. web-page (2).

3. meeting of committee members (2).

   • one-day meeting (1).

   • employ a full committee meeting model (1).

A sample of representative quotes are as follows:

"A "FAQ" [frequently asked questions] prior to the deadline would be helpful on the web or E-mail."

"I would think that the discrepancies should be discussed rather than resolved by a third party."

"I think it would be useful to have some discussion to ensure that the individual members view of the evaluation criteria are comparable."

"We should value graduate students highly enough to make this evaluation process more interactive. It would also remove any tendency for arbitrariness in score assignment since reviewers would have to discuss the merits/demerits of the application in an open forum and justify the scores/evaluation."

"I think a forum, even amongst graduate students applying as well as proposed supervisors, would be an excellent source of information for what the people actually going through the process think of it, as opposed to just reviewers handing down their evaluations on the issues of assessment. It would be interesting to find out how the students thought of the process themselves."

2.6 Additional Comments Provided by Committee Members:

Eleven committee members supplied additional comments. Additional comments were grouped into the four categories:

- concerns (2).

  "I was concerned at how few people applied to this program."

  "I felt that some of the committee members scored applications poorly from smaller institutions because of their perception of the training environment [the committee member conveyed personal knowledge of a particular application]."

  "I am concerned that the social science applicants are rated by physical scientists. Not because they do not understand the science but that the norms, expectations, workloads and outcomes are vastly different in the two areas. MRC's system puts them competitively in the same pot."

- dissatisfaction (2).

  "Although there are certain advantages to a virtual committee, I am not a big fan of it, for reasons outlined earlier. These fellowships can make or break a future career in science and I just don't think we give candidates a fair enough assessment through this virtual process."

  "There was little positive feedback from being a member of the DRA committee. I did not feel very useful."

- suggested improvements (5).

"I wish that you had sent me this form in February, right after I had done it. Actually, I believe that I did send a letter with one page of comments a long time ago. How about one or two conference calls just so that everyone hears the other member's problems and then hopefully the solution is made clear to everyone. This might make scoring more even (if this is a problem) and might eliminate major differences in the scoring method between members."

"It would have been easier, perhaps more beneficial to MRC, to have completed this survey immediately after the review of applications. Some thoughts fade with time! Otherwise, I think that it is a useful follow-up."

"It would have been useful to fill out the form on a website of the MRC.

"Regarding the project description: In the eligibility requirementssection it would be helpful if the eligibility of foreign health professionals were addressed explicitly."

Another improvement might be a better evaluation of the training environment including issues such as frequency of meetings with supervisor during first/second year, departmental advisory/thesis committee meetings, opportunities to present results at meetings
(local; national; international) ,etc.

- general satisfaction (2).

"It was a pleasure and an honour to be involved with this committee.For the most part I think this process works well especially at this level where the evaluation is fairly straightforward. I place importance upon the applicants' ability to write the grant and there does not appear to be any mechanism to ensure that. Can we not have some kind of signed statement where both student and supervisor indicate the degree to which the student and the supervisor contributed to the writing? My sense is that most of the applications were largely scripted by the supervisor! This creates potential bias in cases where the applicant wrote the description of the project."

"Overall, I think the right people were awarded a studentship. Although, I missed the collegiality of a grant review panel, I think it saved the MRC money and worked well by having all the reviewing done "off-site." I liked that "mark alone" did not

determine outcomes. Thank-you for asking me to review these applications.".

Appendix A

Assessment of an Application for an MRC Doctoral Research Award

Area of Assessment				Score	Max	Reviewer's Comments
1. *Characteristics and Abilities of the Candidate*						
Critical thinking	s1	s2	s3		4	
Independence	s1	s2	s3		4	
Perseverance	s1	s2	s3		4	
Originality	s1	s2	s3		4	
Organizational skills	s1	s2	s3		4	
Interest in discovery	s1	s2	s3		4	
Research ability	s1	s2	s3		4	
			Subtotal			
		Adjustment score >			+/- 4	
	Adjusted subtotal (maximum 28)				28	

	Score	Max
2. *Initial Research Activity of the Candidate*		
Publication activity		6
Other research activity		6

	Score	Max
3. *Research Training Environment*		
Scientific activity		4
Research resources		4
Training record		4
Training program for the candidate		8

	Score	Max
4. *Academic Performance of the Candidate*		
Undergraduate academic performance		6
Academic performance as a graduate student		6

s = sponsor's score

Total		72

Appendix B

Guide to the Scoring of Applications for MRC Doctoral Research Awards

Variable Assessed	Information Source	Scoring *All scores should be entered in whole numbers, i.e., no decimals*	Notes to Reviewers
1. Characteristics and Abilities of the Candidate			
Critical thinking **Independence** **Perseverance** **Originality** **Organizational skills** **Interest in discovery** **Research ability**	**Reports from the candidate's three sponsors**	**Characteristics and abilities of candidates have been scored by their three sponsors on a scale of 1 to 4.** **The sponsors' individual sores and the mean scores for each item are pre-printed on the assessment form provided by MRC.** **Reviewers should read the sponsors' narratives about each candidate and enter an adjustment score if appropriate. (See notes, next column)**	**Assess the extent to which the box scores and narratives are consistent.** **If on balance the narratives do not seem to substantiate the box scores, adjust the subtotal downwards (-1, -2, -3 or -4)** **Conversely, if on balance the narratives produce a more favourable impression than the box scores, adjust the subtotal upwards (+1, +2, +3 or +4). Note, however, that the adjusted subtotal must not exceed 28 points.**
2. Initial Research Activity of the Candidate			
Publication activity **(review the list of articles and other publications produced by the candidate)**	**CV module completed by the candidate**	**1** **Very low** **2** **Low** **3** **Moderate** **4** **High** **5** **Very high** **6** **Outstanding**	**Assess the publication activity of the candidate relative to your expectations of someone with their academic experience.** **Consider: breadth of science covered, the frequency of publication and the scientific impact of the journals involved.** **In considering the candidate's input to the publications, take into account the number of coauthors for each paper and the prominence of the candidate's name on the list of authors.**
Other research activity **(review information on presentations, research prizes and other indicators of the candidate's research productivity)**	**as above**	**1** **Very low** **2** **Low** **3** **Moderate** **4** **High** **5** **Very high** **6** **Outstanding**	**Assess other research activity of the candidate relative to your expectations of someone with their academic experience.** **Consider: breadth of science covered, size and importance of meetings involved, frequency of conference presentations and research honours or awards.**
3. Research Training Environment			
Scientific activity **(review the research supervisor's publication record, significant contributions to research, and honours or awards)**	**CV module completed by the research supervisor**	**1** **Moderate** **2** **Good** **3** **Very good** **4** **Excellent**	**In assessing whether the scientific environment is one in which a doctoral student will be inspired and challenged, focus on the scientific productivity and impact of the research supervisor as indicated by publications, presentations and research prizes or awards.** **Bear in mind that publication activity patterns vary among health science disciplines.**

Variable Assessed	Source	Scoring		Notes for Reviewer
Research resources (review the research resources available in the laboratory in which the candidate will train)	**CV module completed by the research supervisor**	1 2 3 4	Moderate Good Very good Excellent	In assessing the extent to which the level of research resources in the training environment will enhance the scientific development of the candidate, focus on the adequacy of peer-reviewed research funding secured by the research supervisor. Bear in mind that availability of funding varies among health science disciplines.
Training record (review the supervisory experience of the researcher under whom the candidate proposes to pursue doctoral studies)	**as above**	1 2 3 4	Moderate Good Very good Excellent	Consider the number of masters students, doctoral students and postdoctoral fellows that the research supervisor has trained or is currently training. In assessing the likelihood that students will be inspired to continue on in research, consider any information on the subsequent positions of persons who trained with the research supervisor.
Training program for the candidate (review the candidate's training expectations and proposed doctoral research program, including project, resources available and planned non-research activities)	**Training module completed by the candidate**	1 2 3 4 5 6 7 8	Exceptionally weak Very weak Weak Moderate Fair Good Very good Excellent	Most candidates will be conducting research outside your scientific specialty. From a non-specialist's perspective, assess the intellectual challenge and excitement of the research in which the candidate will be involved. Consider the extent to which the training program appears to fit with the candidate's training goals and the resources available for the candidate's project.

4. Academic Performance of the Candidate

Variable Assessed	Source	Scoring		Notes for Reviewer
Undergraduate academic performance	**Academic transcripts of the candidate**	1 2 3 4 5 6	Very weak Weak Moderate Strong Very strong Exceptional	**Consider:** - type of program and courses pursued - course load - grades obtained - relative standing (if available) - overall average - trend (give credit for a steadily improving or consistently good performance)
Academic performance as a graduate student	**as above**	1 2 3 4 5 6	Very weak Weak Moderate Strong Very strong Exceptional	as above

Appendix C

Sponsor Report Form

Name of Candidate

Name of Sponsor

Position/Department/Institution

Number of years that the sponsor has known the candidate	Capacity in which the sponsor has known the candidate

This report will be used by the Medical Research Council of Canada (MRC) when assessing the candidate's application for an award in support of doctoral study in the health sciences. Note that according to the Canadian Privacy Act, if the candidate specifically requests a copy of this report, the Council must make it available.

There are two parts to the report. In Part A, sponsors will score the candidate's performance on seven dimensions. In Part B, they will explain the reason for each score. Both Part A and Part **B of the report must be completed.**

The report should be sent to the candidate in a sealed envelope with the sponsor's signature across the seal. Sponsors should bear in mind that the candidate will need to assemble a complete application package, including this report, by the competition deadline of October 15. The Council will not consider late or incomplete applications.

Part A: RATING FORM

Please indicate your perception of the candidate on each dimension by circling the box score which best reflects your view of the candidate's performance. A circling of two adjacent scores is acceptable and will be interpreted as the mean of the two (e.g., circling 3 and 4 generates a score of 3.5). You will be required to give specific examples of the candidate's behaviour to support these ratings in Part B of the report.

Critical thinking - judicious evaluation of all information, regardless of its source				
1 rarely exhibits critical thinking	2 sometimes exhibits critical thinking	3 often exhibits critical thinking	4 always exhibits critical thinking	*unable to judge*

Independence - pursuit of knowledge or taking of action on own initiative, seeking guidance only when appropriate				
1 rarely exhibits independence	2 sometimes exhibits independence	3 often exhibits independence	4 always exhibits independence	*unable to judge*

Perseverance - determined persistence in pursuit of goals despite obstacles or discouragement				
1 rarely exhibits perseverance	2 sometimes exhibits perseverance	3 often exhibits perseverance	4 always exhibits perseverance	*unable to judge*

Originality - imagination or ingenuity in problem solving				
1 rarely exhibits originality	2 sometimes exhibits originality	3 often exhibits originality	4 always exhibits originality	*unable to judge*

Organizational skills - systematic, careful planning and coordination of activities				
1 rarely exhibits organizational skills	2 sometimes exhibits organizational skills	3 often exhibits organizational skills	4 always exhibits organizational skills	*unable to judge*

Interest in discovery - an enquiring mind and a strong desire to pursue new knowledge				
1 rarely exhibits interest in discovery	2 sometimes exhibits interest in discovery	3 often exhibits interest in discovery	4 always exhibits interest in discovery	*unable to judge*

Research ability - a natural talent or acquired proficiency for scientific investigation				
1 rarely exhibits research ability	2 sometimes exhibits research ability	3 often exhibits research ability	4 always exhibits research ability	*unable to judge*

Name of candidate

Part B: EXPLANATION OF RATINGS

To help explain your ratings, please provide specific examples of the candidate's behaviour with respect to each characteristic in Part A: Critical thinking; Independence; Perseverance; Originality; Organizational skills; Interest in discovery; and Research ability.

Signature of Sponsor	Date

(If you prefer to print Part B on plain paper, please ensure that the name of the candidate appears at the top of the page and your signature, and the date, at the end. Part B must not exceed two pages.)