

54. Enriching program evaluation through insiders' opinions: An American university graduate program¹

Kadir Vefa TEZEL²

APA: Tezel, K. V. (2021). Enriching program evaluation through insiders' opinions: An American University graduate program. *RumeliDE Dil ve Edebiyat Arařtırmaları Dergisi*, (23), 861-874. DOI: 10.29000/rumelide.949718.

Abstract

Assessing the quality of education is a vital stipulation for all graduate programs that aim to secure and bolster their current rank among similar programs. Program evaluation is usually carried out by parties not directly related to a program. In other words, outsiders evaluate programs. Insiders who have first-hand experiences and impressions formed on those experiences are typically excluded from the evaluation process. Inclusion of the perspectives of insiders into the gathered data in program evaluation, however, would certainly add depth and enrich reports produced at the end of program evaluation studies. Another important point is the limitation of traditional curriculum evaluation studies due to the fact that they do only one of the four types of evaluations that exist in the literature. Including the views of insiders in the evaluation process and choosing a data collection tool that will enable multiple types of evaluation to be done in a single effort will also contribute to a more comprehensive assessment of graduate programs. In order to carry out such a program evaluation, this study conducted an evaluation based on the impressions of insiders about the aspects related to the quality of education provided in an American university graduate program. The participants were all faculty members and students in the program. The results showed that both groups expressed positive opinions about the majority of the dimensions directly related to the quality of the education provided in the program. It was also determined that there were areas that were not primarily related to the quality of education which needed improvement, but that most of those areas would become areas of high satisfaction with little effort.

Keywords: Program evaluation, graduate education program, survey research, questionnaire

Programın parçası kiřilerin grřleriyle zenginleřtirilmiř program deęerlendirmesi: Bir Amerikan niversitesi yksek lisans programı*

z

Eęitimin kalitesini deęerlendirmek, benzer programlar arasında mevcut sıralarını korumayı ve gçlendirmeyi amalayan tm lisansst programlar iin hayati bir kořuldur. Program deęerlendirmesi genellikle bir programla doęrudan iliřkili olmayan taraflarca yapılır. Bir bařka deyiřle, program deęerlendirmesini dıřarıdan gelen kiřiler yapar. Programın parçası olan ve onunla ilgili ilk elden deneyimlere ve bunlara iliřkin izlenimlere sahip olan insanlar ise genellikle deęerlendirme srecinin dıřında tutulurlar. Oysa program deęerlendirmesinde toplanan verilere programın parçası olan kiřilerin bakıř aıllarının dhil edilmesi, program deęerlendirme alıřmalarının sonunda retilen raporlara kesinlikle derinlik katacak ve onları zenginleřtirecektir.

¹ Bu alıřma yazarının doktora tezinden retilmiřtir.

² Dr. ęr. yesi, Bolu Abant İzzet Baysal Universitesi, Eęitim Fakltesi, Yabancı Diller Eęitimi Blm (Bolu, Trkiye), vefa2000@gmail.com, ORCID ID: 0000-0002-2636-1221 [Arařtırma makalesi, Makale kayıt tarihi: 16.04.2021-kabul tarihi: 20.06.2021; DOI: 10.29000/rumelide.949718]

Bir başka önemli konu da geleneksel program değerlendirme çalışmalarının alan yazında var olan dört tür değerlendirmeden sadece bir türünü yapmasından kaynaklanan sınırlılıklardır. Bir programın parçası olan insanların görüşlerini de içeren ve birden çok değerlendirmenin tek seferde yapılmasını mümkün kılacak bir veri toplama aracı seçmek, lisansüstü programların daha kapsamlı bir değerlendirilmesine önemli katkılarda bulunacaktır. Bu tür bir program değerlendirmesini gerçekleştirmek amacıyla, bu çalışmada bir Amerikan üniversitesinin yüksek lisans programında verilen eğitimin kalitesi ile ilgili yönleri hakkında programın parçası olan insanların izlenimlerine dayalı bir değerlendirme çalışması gerçekleştirilmiştir. Katılımcılar programdaki tüm öğretim üyeleri ve öğrencilerdir. Sonuçlar, her iki grubun da programda verilen eğitimin kalitesiyle doğrudan ilgili boyutların çoğu hakkında olumlu görüş bildirdiklerini göstermiştir. Eğitim kalitesiyle öncelikli olarak ilgili olmayan ancak iyileştirilmesi gereken alanların da olduğu, ancak bu alanların çoğunun çok az bir çabayla yüksek memnuniyet duyulan alanlara dönüşeceği de belirlenmiştir.

Anahtar kelimeler: Program değerlendirme, lisansüstü eğitim programı, tarama araştırması, anket

1. Introduction

Educational institutions strive to enhance the quality of education they give because they gain recognition and establish a good reputation for themselves in return. Quality is not a dimension that offers itself readily to assessment in education. It is assessed through a process of collecting and assessing relevant data to make evaluative judgments about a program referred to as program evaluation in which data are collected, analyzed, and interpreted to make judgments as to whether to continue a program as it is, modify it, or terminate it (Craven, 1980). Programs are evaluated to determine the quality of education given in them (Herman et al., 1987) and necessary improvements are subsequently made to improve educational quality.

Program evaluation is generally carried out by external parties such as accreditation agencies who are not parts of the program. These people or groups focus on aspects of programs like the number of students, the amount of research conducted by faculty members, and the program's regional and national ratings in their reports. Evaluation reports prepared by those parties are limited in terms of comprehensiveness for three main reasons: first, people who prepare them are outsiders who are invited and sometimes financially compensated to conduct the evaluation. These people do not possess first-hand experiences or opinions about the program. Using their expertise, they make use of the available data and a set of criteria or standards to judge how well the program fares. Even though such reports are without a doubt valuable for stakeholders to make decisions, they lack crucial data, the inclusion of which would provide information that would certainly surpass the extent of traditional evaluations. Those crucial data would significantly contribute to the comprehensiveness of program evaluation reports because they would consist of the opinions of people who are the parties that are directly involved in those programs; these people are faculty and students. They are the insiders. No other party can make more truthful value judgments that are inherently related to the internal aspects of a program than these parties. The inclusion of the opinions of these parties is essential to make healthier decisions to develop or improve a program.

Second, program evaluations are limited in scope. In the usual approach to program evaluation, only one kind of evaluation is usually conducted. Brown (1995, pp. 219-224) states that the four approaches

in doing language program evaluation are product-oriented, process-oriented, static characteristic, and decision facilitation approaches. The limitations of these approaches lie in the fact that product-oriented evaluation is summative and examines how well the goals of a program are achieved. Process-oriented evaluation is formative and takes a snapshot of current practices in an existing program to determine whether making changes are necessary. Static characteristic evaluation examines how effective a program is in terms of the facilities it has to offer such as library holdings, and the like. Decision facilitation approach is used to furnish information for decision-makers to make informed decisions.

This study was designed to show that it is possible to collect data to do more than one type of evaluation in a single evaluative effort. Thus, it is product-oriented as it aimed to find out to what extent the program's goals were achieved. It is process-oriented as it aimed to collect information that was useful in determining the changes to be made to improve program. It is static as it collected information on the existing facilities and faculty. It also has decision facilitation features because the collected data were reliable and could readily be used to aid administrators to make decisions.

Third, fiscal cost is a significant concern in evaluative studies. Keeping expenditures at reasonable levels is one of the desired goals for evaluation studies. In routine evaluation studies, budgets have to be formed to cover the expenses that arise in different phases of the evaluation. Unlike the usual practice, this study was conducted by one evaluator to keep the fiscal expenses at the lowest possible limit.

2. Review of literature

It is customary for departments to conduct evaluations to assess the current level of satisfaction with the program and identify the areas that need improvement. As such reports contain sensitive information, departments use these reports for program improvement purposes and do not make them public because they do not want their weaknesses to be known by others. For those interested in the subject, this course of action results in finding only a limited number of evaluation studies in the professional literature. This is a serious limitation because researchers who look for actual evaluation reports end up finding very few studies but numerous books and articles on how to conduct program evaluation. Regarding graduate program evaluation Clark (1976, pp. 85-93) remarks that the three common ways are reputational rating, accreditation, and individual program evaluation studies.

2.1.Reputational rating approach

Pioneered by Hughes (1925), the term "reputational rating" refers to the evaluation of a program by field experts. According to Conrad and Blackburn (1985), a criterion is selected prior to conducting a reputational rating study. Faculty quality is the primarily considered criterion. Experts are then asked to rate a program using that criterion. Reputational ratings are generally conducted on a wider scale in an academic field and a list of universities ranked according to the criterion is created afterwards. Consequently, top-ranked universities in the prepared list receive more financial aid, hire well-known scholars as faculty members more easily, and become the primary choice of students. Their graduates also find better jobs in the job market.

The reputational rating approach is important in program evaluation in that it is the first in using faculty quality as the evaluation criterion by directly obtaining the opinions of insiders in program

evaluation. Hughes' approach received criticism because it only sought the opinions of faculty members who were the staff of universities such as Harvard and Yale which were already considered high-quality schools by the public and did not include the opinions of a greater number of scholars who worked in all other universities. Hughes' approach was improved by Keniston who made the approach credible by using experts as raters in assessing scientific, scholarly, and artistic quality in educational programs (Diamond & Graham, 2000, p. 22).

A truly large-scale reputational study of 2.699 graduate programs was conducted in 1982. This study went beyond the usual faculty quality criterion and included criteria such as the assessment of university libraries, graduate student profiles, and numbers of faculty members (Webster, 1983) to widen the typical scope of such studies. In their present format, reputational ratings are still used to rank graduate programs in the United States.

2.2. Accreditation

Accreditation is another way of assessing quality in graduate programs in which the quality of education in a program is reviewed by a competent organization, using a set of quality standards (Hamalainen, 2003, p.298). Accreditation is beneficial for schools because it helps them determine what they need to do in order to improve and maintain quality in education (Millard, 1983) and ensures the allocation of funds (Semrow, 1981). Accredited schools and their programs are favored by the public because diplomas from those schools provide better-paying jobs and living conditions.

Accreditation is criticized because accreditation organizations employ quantifiable criteria such as student enrolment figures which inform stakeholders as to what degree something exists if it does, but they fail to explain the reason/s for their presence or absence (Millard, 1983). The lack of a common format that includes a set of standard procedures is also a problem in accreditation studies (Hamalainen, 2003, pp. 292-293).

To overcome this problem, Stufflebeam (2001, p. 72) recommends the use of evaluation checklists. These lists are useful in guiding evaluations and standardizing them. Some well-known checklists are the "Qualitative Evaluation Checklist" (Patton, 2003) and the "Key Evaluation Checklist" (Scriven, 2005) which are available with an array of other evaluation checklists free of charge at <https://wmich.edu/evaluation/checklists>.

2.3. Individual program evaluations

Individual program evaluations are conducted for departmental evaluation purposes and as such their results are not commonly made public as they consist of information on the strengths and weaknesses of programs which administrators keep for themselves. In a rare exception to this tendency, Fradd and Lee (1997) shared the results of an individual program evaluation conducted to identify the aspects of the program that needed to be modified to address the needs of the students as best as possible.

Another rare individual program evaluation study that publicized its results was conducted by Kayla, Wheelless, and Howard (1981) who decided to develop a valid and reliable questionnaire that would capture the opinions of students doing graduate studies in the United States. The researchers developed a questionnaire that consisted of 39 items to collect information on six crucial aspects of any graduate program. This study is important in the sense that the instrument was reliable and could be

used confidently to capture the opinions of participants on six different aspects instead of a single criterion.

As stated in the introduction, this study aims to present a comprehensive, practical, and financially feasible format to do program evaluation by gathering the impressions of insiders to assess the quality of a graduate program. To that end, it used a set of questionnaires specifically designed to collect data primarily on the quality of education given in a graduate program in addition to other aspects not directly related to educational quality with the intention to do a more comprehensive program evaluation than the limited single-focus evaluation studies and sought answers to the following questions:

1. What are the opinions of the faculty members and students about the common aspects of the program?
2. What are the opinions of the faculty on the faculty-specific aspects of the program?
3. What are the opinions of the students on the student-specific aspects of the program?

3. Method

3.1. Research design and data collection tool

Qualitative and quantitative designs are used in evaluation studies (Worthen and Sanders, 1987; Patton, 1987, 1990). Dabbs (1982) states that qualitative research refers to meanings, concepts, definitions, characteristics, metaphors, symbols, and descriptions of things whereas quantitative research refers to counts and measures of things. According to King et al. (1987), the conventional approach to program evaluation is quantitative. In this approach, information about an educational program is expressed in numerical terms. Interpretations based on quantitative data inform stakeholders on the absence or presence of a feature in a program and its extent. Sonquist and Dunkelberg (1977) remark that in program evaluation the survey method is used.

This study, which aims to evaluate educational quality in a graduate program, collected data from multiple insider parties, staying in the quantitative paradigm and using the survey method. It employed a quantitative data collection survey tool named the Graduate Program Self-Assessment (GPSA) questionnaires which were designed to collect quantifiable data which were of primarily qualitative nature. According to Nunan (1992, pp. 140-142), the most common types of surveys which are convenient tools for recording conditions and people's attitudes at a single point in time are questionnaires and interviews.

Using the survey method, the data were collected at an American university. To protect the identity of the school and not hurt its reputation, the name of the university will be kept confidential. The GPSA questionnaires allow data triangulation which was done by cross-checking the data obtained from the groups to strengthen the findings and allow the collected data to be used to do the four types of evaluation stated by Brown (1995). There are separate questionnaire booklets for each group which include identical and group-specific items. The validity of the questionnaires was established by Clark (1974); Clark et al. (1976), and Rowshan (1988). A reliability study was conducted by Clark et al. (1976) as well.

3.2. Data collection in the GPSA questionnaires

In order to obtain a comprehensive snapshot of the quality of education in the program, all participants were included in the sample. The participants were five faculty members and sixteen students who were taking courses in the program at the time of data collection.

All five faculty members filled out and returned the questionnaires with a return rate of 100%. Of the distributed sixteen student questionnaires, ten were returned which represented a 63% return rate. As any return rate over 50% is acceptable (Glastonbury & MacKean, 1991, p. 235), these percentages indicated satisfactory return rates.

3.3. Data analysis

The questionnaires were sent to the Educational Testing Service (ETS) for data analysis. ETS conducted descriptive analyses and the results were mailed to the researcher. The participants expressed their opinions on 16 aspects of the program by responding to a total of 105 items. 47 of those sought the opinions of the two groups on the eight common aspects of the program. The faculty responded to 28 additional items, while the number of additional items was 30 for the student group.

Five different types of response choices are given to respondents to answer the items in the questionnaires. These types are given in the order of their frequency of usage along with the research questions they are used in. Type 1 offers four choices which are (1) disagree strongly, (2) disagree with reservations, (3) agree with reservations, and (4) agree strongly. This type was used in answering all research questions in this study. Type 2 offers four choices which are (1) poor, (2) fair, (3) good, and (4) excellent. This type was used in obtaining the participants' impressions on the research questions 1 and 3. Type 3 offers two choices which are (1) yes and (2) no. It was used in answering the research questions 2 and 3. Type 4 offers five choices which are 1) want much more emphasis, (2) want some more emphasis, (3) want the same emphasis, (4) want some less emphasis, and (5) want much more emphasis. This type was used in answering the research question 1. Type 5 offers three choices which are (1) never, (2) occasionally, and (3) frequently. It was used in answering the research question 2.

4. Findings

In the descriptive analyses by ETS, mean scores were calculated to find the similarities and differences between the opinions of the two groups. A four-point scale is used in the interpretation of the mean scores of the groups. Those points and their referent explanations are (1) poor, (2) fair, (3) good, and (4) excellent. According to Clark et al. (1976, p. 54), a mean score of 3.00 or above is "very good". Mean scores below that number indicate areas that require attention.

Having furnished this information essential for the interpretation of the findings, the answers to the research questions will be presented in the tabular form first, followed by necessary explanations.

Research question 1: Faculty and student opinions on the common aspects of the program

To answer this research question, faculty and student opinions were elicited on eight common aspects of the program. Those aspects and the mean scores of the groups are presented in Table 1.

Aspect of program	Faculty mean score	Student mean score
1. Environment for learning	3,00	3,36
2. Scholarly excellence	3,10	3,30
3. Quality of teaching	3,35	3,28
4. Faculty concern for students	3,05	3,17
5. Curriculum	2,83	2,71
6. Departmental procedures	2,65	2,88
7. Available resources	2,17	2,71
8. Faculty and student perceptions on the program purposes	varying opinions	varying opinions

Table 1. Faculty and student opinions on eight common aspects of the program

Three types of response choices were given to both groups in stating their opinions on those common aspects of the program. Type 1 was used to elicit the opinions of both groups on aspects 1, 2, 3, 4, and 6 of the program. Type 2 was used for aspects 5 and 7. Type 3 was used to obtain the opinions of the groups on aspect 8 of the program.

Aspect 1: Environment for learning

Five items were used to obtain the opinions of faculty members and students on the learning environment aspect of the program. The items were used to elicit the groups' opinions on how supportive the program was for students, whether there was mutual respect between faculty and students, how helpful students were towards each other, and whether the department was open to novelty and different opinions. The faculty mean was (M= 3.00) and the student mean was (M= 3.36). These mean scores indicated that both parties rated this aspect of the program as "very good".

Aspect 2: Scholarly excellence

Six items were used to elicit the opinions of both parties on three main points which were faculty excellence, student capabilities, and the extent of intellectual stimulation. The means showed that both groups rated this aspect as "very good". Faculty mean was (M=3.10) and student mean was (M=3.30).

Aspect 3: Quality of teaching

Six items were used to obtain the opinions of both groups on two major points which were openness of faculty members to new ideas and their helpfulness to students, and students' opinions on the teaching methods and grading procedures faculty members used and faculty members' levels of preparation for classes. The means of both parties were "very good". Faculty mean was (M= 3.35) and student mean was 3.28.

Aspect 4: Faculty concern for students

In this category four items were used to elicit the opinions of the parties on issues such as the extent of faculty concern in students' professional development, faculty members' making themselves accessible to students, and faculty members' awareness of students' needs, concerns, and suggestions. The means of the groups were once again "very good"; faculty mean was (M=3.05) and student mean was (M=3.17).

Aspect 5: Curriculum

Six items were used to elicit the opinions of both parties on curriculum related issues such as the variety of the offered courses, the flexibility and opportunities to pursue available to carry out individual projects in the program, and interaction with related departments.

The mean scores of both parties were below 3.00. Faculty mean was (M=2.83) and student mean was (M=2.71). As the mean scores of both parties show, curriculum was an area of concern for both parties: no member of the faculty believed that the program's interaction with other programs or related disciplines was excellent. Excellent rating on this aspect of the program was given by a very small percentage of students (10%). As for the extent of the program's interaction with other programs, different opinions were expressed by the groups: the percentage of *good* rating was 60% in the faculty group while it was 30% in the student group. *Fair* rating by faculty was 20% and by students it was 10%. Poor rating percentages were also considerable. Faculty percentage was 20% and that of the students was 10%. The percentage of students who did not express their opinions on this aspect was high by 40%.

Aspect 6: Departmental procedures

Eight items were used to obtain the opinions of both groups on departmental policies and procedures. Once again, the means of both parties were below 3.00. Faculty mean was (M=2.65) and student mean was (M=2.88). The analysis showed that finding employment for students following graduation, student-centered advising, student participation in departmental decisions, and evaluation of students by the department were dimensions not given priority in the program.

Aspect 7: Available resources

Seven items were used to elicit the opinions of the groups on the availability of university facilities and the adequacy of resources in the program. This is another aspect where the means of both parties indicated problems. Faculty mean was (M=2.17) and student mean was (M=2.71). Even though these mean scores were low, it is necessary to state that the data on this aspect merely provided information on the sufficiency of available resources, not on a quality-related dimension of the education given in the program.

Aspect 8: Faculty and student perceptions on program purposes

The participants responded to five items to express their opinions. Their opinions are presented in percentages in Table 2 below.

	Faculty (degree of emphasis)					Students (degree of emphasis)				
	much more	some more	same	some less	none	much more	some more	same	some less	none
1.Preparing scholars and researchers	20	40	40	0	0	10	0	60	20	0
2.Preparing teachers	0	0	60	40	0	0	20	50	10	0
3.Preparing other practitioners	0	80	20	0	0	20	0	30	0	0
4.Preparing students	20	20	40	20	0	0	40	40	10	0

for more advanced study

5.Providing personal enrichment	20	40	40	0	0	30	20	30	10	0
---------------------------------	----	----	----	---	---	----	----	----	----	---

Table 2. Faculty and student perceptions on the program purposes

Program purpose 1: Preparing scholars and researchers

Less than half the faculty (40%) and more than half the students (60%) were satisfied with the attention currently given to this goal of the program. 60% of the faculty stated that more emphasis on this purpose would be appropriate; of those faculty members, 40% expressed the need for some more emphasis and 20% wanted much more emphasis as was the case by 10% of the students. Students were the only party that demanded less emphasis. This opinion was voiced by 20% of the students.

Program purpose 2: Preparing teachers

The percentages of the groups that wanted the same emphasis on this purpose to continue were 60% for faculty and 50% for students. Students were the only group that wanted some more emphasis by 20%. As for lesser emphasis, both faculty and students expressed this opinion by 40% and 10% respectively.

Program purpose 3: Preparing other practitioners

A considerable percentage of faculty (80%) expressed the need for more attention in this respect. 20% of students wanted much more emphasis. Maintaining the same emphasis was expressed faculty and students at percentages of 20% and 30% respectively. 50% of students did not respond to this item.

Program purpose 4: Preparing students for more advanced study

An equal percentage of members in both groups (40%) expressed the opinion that the current emphasis should continue. The percentages of members in both groups that wanted more emphasis were 20% for faculty and 40% for students. The percentage of faculty that believed much more emphasis should be placed on this program purpose was 20%. There were also members in both groups who stated that lesser emphasis should be given to this purpose; the percentages were 20% for faculty and 10% for students.

Program purpose 5: Providing personal enrichment

Less than half the members of both groups, faculty (40%) and students (30%), stated that no changes needed to be made in this regard. 40% of faculty and 20% of students believed some more emphasis would be better. The percentages of members of both groups who believed much more emphasis on this purpose would be appropriate were 20% for faculty and 30% for students. 10% of students believed placing some less emphasis on this purpose would be adequate.

Research question 2: Faculty opinions on the faculty-specific aspects of the program

To answer this research question, faculty members' opinions were elicited on four aspects of the program. The aspects and faculty mean scores are given in Table 3.

Aspect of program	Faculty mean score
9. Faculty work environment	2,87
10. Faculty program involvement	2,30
11. Faculty research activities	33%
12. Faculty professional activities	48%

Table 3. Faculty opinions on the faculty-specific aspects of the program

Three different response choices were given to faculty members to state their opinions on the faculty-specific aspects of the program. For aspect 9, Type 1, for aspect 10, Type 5, and for aspects 11 and 12, Type 3 were given.

Aspect 9: Faculty work environment

Nine items were used to elicit the opinions of faculty on this faculty-specific aspect of the program which asked faculty members to share their views on their satisfaction with the objectives and procedures, academic freedom, participation in decisions, and personal relationships with colleagues in the department where the program was offered. The low faculty mean score of (M=2.87) indicated that faculty members did not find the work environment conducive to performing at the level they desired.

Aspect 10: Faculty extent of program involvement

Eight items were used to elicit the perspectives of faculty on this aspect of the program to determine the extent they were involved in the program such as teaching courses, participating in departmental examinations, and directing independent studies and theses. ETS's evaluation of this aspect is based on a scale in which 3.00 is the highest possible mean. The faculty mean (M=2.30) indicated an overall satisfaction by group members.

Aspect 11: Faculty research activities

Six questions were directed at faculty members to determine the extent of research activities they were engaged in such as awards they had received, editorial and refereeing responsibilities in scholarly journals, and financial support they had received for their research or other scholarly activities. Faculty mean was 33%. This low mean indicated that the group members were not actively involved in research activities besides teaching courses.

Aspect 12: Faculty professional activities

Five questions were asked to determine the extent faculty members were conducting professional activities such as serving on any national or advisory committees, serving as officials in local or national professional organizations, and being the recipients of awards for outstanding professional performance. The mean was 48%. This low mean showed that faculty members were performing professional activities at a slightly lower level than the median 50%.

Research question 3: Student opinions on student-specific aspects of the program

To answer this research question, students' opinions were sought on four aspects of the program which are given below in Table 4 along with student mean scores.

Aspect of program	Student mean score
13. Student satisfaction with program	3,32
14. Assistantship and internship experiences	-
15. Resource accessibility	2,38
16. Student accomplishment in the last twelve months	37%

Table 4. Student opinions on the student-specific aspects of the program

Three different response choices were given to students to elicit their views on the four student-specific aspects of the program. For aspect 13, Type 1, for aspects 14 and 15, Type 2, and for aspect 16, Type 3 response choices were given.

Aspect 13: Student satisfaction with the program

Four items were used to obtain student opinions on their satisfaction with the program, willingness to recommend the program to others, knowledge they had acquired in the program, and the level of professional readiness. The mean was ($M=3.32$). This score showed that student satisfaction with the program was high.

Aspect 14: Assistantship and internship experiences

Six questions were directed at the students to elicit their opinions on this aspect. Three students responded to the questions on this aspect whereas ETS required at least five respondents to conduct data analysis. Thus, no analysis was conducted due to the low number of respondents.

Aspect 15: Resource accessibility

Five items were used to elicit student opinions on the accessibility of services offered to students by the university. Student mean was ($M=2.38$). It is necessary to emphasize that this low score does not indicate a dimension directly related to the quality of education given in the program.

Aspect 16: Student accomplishment in the last twelve months

Fifteen items were used to determine the individual professional activities students were engaged in prior to the administration of the questionnaire.

The mean was 37%. This low mean does not necessarily indicate a problem: As students were in the process of honing their professional knowledge and skills, these results may be interpreted as indicators of the beginning of their professional lifelong journey. Table 5 presents all items and students' responses to them.

	Yes	No	Omit
1. Attended meeting of a scholarly or professional society	100		

2. (Co)authored a paper accepted for presentation at a scholarly/professional meeting	10	90	
3. (Co)authored a paper submitted for publication in a scholarly/professional journal		90	10
4. Demonstrated artistic skills or products in a public performance or exhibit	20	80	
5. Prepared detailed proposal or plan for master's thesis or other major project	30	70	
6. Carried out an independent research or creative project	90	10	
7. Cooperated in research or creative project with a student or faculty member	50	40	10
8. Held a fellowship, training grant, or scholarship	10	90	
9. Developed professional skills thru clinical, field work, internship experiences	30	70	
10. Talked with professionals in field about other graduate programs or career plans	70	30	
11. Pursued independent reading or practice in the field beyond courses	90	10	
12. Operated independent enterprise or business			100
13. Won a prize or an award for a product or an activity related to field	10	90	
14. Served on a department or university-wide committee	20	80	
15. Participated in department or program planning (e.g. review of the curriculum)	20	80	

Table 5. Student accomplishment in the last twelve months

5. Discussion and conclusion

Programs may be evaluated by employing qualitative or quantitative designs. In this study, the traditional quantitative approach was used. However, to elicit the opinions of insiders on aspects that were directly related to the quality of education given in the program to present a more in-depth picture and carry out program evaluation in a time-saving and cost-effective manner using a single evaluator, this study used the GPSA questionnaires which are quantitative by nature to elicit insiders' views on the quality of education given in an MA program in an American university. Reliable and valid questionnaires that collect quantifiable data on the qualitative aspects of a program not only allow the inclusion of insiders' opinions in the evaluation but also allow decision-makers to compare them as well. Therefore, the use of such questionnaires would be beneficial to more accurately assess the quality of education in a program by identifying its strengths and weaknesses.

Collecting data from multiple insider parties, namely, faculty and students, allowed data triangulation to be made, provided a basis for the researcher to make comparisons between the two groups, and strengthened the reliability of the findings in this study.

Faculty and students are the real insiders and no other party can provide more essentially significant information to assess the quality of education in a program. Their opinions are crucial and have to be obtained because data obtained from insiders give first-hand information on the actual implementation of a program. External agency evaluations, on the other hand, rely solely on quantitative data to make evaluative judgments about a program. If evaluations are based on the opinions of all parties involved in a program, the obtained data will stand out in terms of direct relevance and guide the decisions to be made better. For programs that intend to carry out evaluations through external evaluation agencies, using questionnaires such as the one used in this study and combining the findings with traditional types of evaluation reports will increase the usefulness of the evaluations and enable decision-makers to make more informed decisions regarding the termination or modification of existing practices and introduction of new ones in the light of the more comprehensive information obtained.

For program aspects 1, 2, 3, 4, and 8, faculty and student mean scores indicated that both parties' evaluative opinions on program aspects related to the quality of education were very good. Data analysis revealed areas of concern as well. The low mean values expressed by both parties for aspects 5, 6, 7, 9, 11, 12, and 15 of the program initially attracted attention as potentially important problem areas. However, the examination of those means showed that they were not about parts of the program that were directly related to the educational quality in the program. To be precise, those scores indicated causes of discomfort and inconvenience to the insiders. Providing solutions to those problems would allow the program to continue in a problem-free way. For program aspect 16, the responses to the items indicated that the low means were not directly related to the quality of education in the department either.

Regarding the insiders' dissatisfaction with some aspects related to the quality of education in the program, of the seventy-five statements faculty members responded to, thirty items (40%) were rated below 3.00. However, eighteen of those items were in the 2.51-3.00 range. This means 60% of the items in which faculty expressed dissatisfaction was very close to becoming "very good". As for the dissatisfaction of students with the program, the analysis showed that of the seventy-seven items they responded to, the number of items in which dissatisfaction was expressed was eighteen (23%). Nine of those items (50%) were also very close to becoming "very good". These results indicate that major modifications were not necessary to improve scores in those items.

Departments do not commonly make the evaluation results of their programs public because weaknesses reported in those reports impact student enrolment and the programs' reputation negatively. It is possible to make evaluation results public as this study has done without negatively affecting the reputation of programs by not disclosing the identities of universities, their programs, and the courses offered in them. Even though keeping evaluation reports confidential is a perfectly understandable course of action, the lack of a substantial number of such reports causes uncertainty on the part of educators who want to identify the strengths and weaknesses of their programs locally by themselves. This scarcity creates demand for external bodies for program evaluations and keeps the demand alive. Not many departments are financially capable of hiring external agencies to conduct evaluations for their programs. It is the hope of this researcher that program evaluation studies such as this one will increase in number in the future because this would allow interested parties to locate and access concrete evaluation reports instead of a vast array of books or chapters on how program evaluation should be conducted.

References

- Brown, J. D. (1995). *The elements of language curriculum: A systematic approach to program development*. Heinle & Heinle.
- Clark, M. J. (1974). *The assessment of quality in Ph.D. programs: A preliminary report on judgments by graduate deans* (GRE Board Research Report No.72. 70a). Educational Testing Service.
- Clark, M. J., Hartnett, R. T., & Baird, L. L. (1976). *Assessing dimensions of quality in doctoral education: A technical report of a national study in three fields*. Educational Testing Service.
- Clark, M. J. (1976). The meaning of quality in graduate and professional education. In J. Katz & R. T. Hartnett (Eds.), *Scholars in the making: The development of graduate and professional students*. (85-102). Ballinger Publishing Co.
- Conrad, C. F., & Blackburn, R. T. (1985). Program quality in higher education: A review and critique of literature and research. In John C. Smart (Ed.), *Higher Education: Handbook of Theory and Research*, Vol. 1, (283-308). Agathon Press.

- Craven, E. G. (1980). A concluding perspective. *New directions for institutional research*, 27 (3), 105-116.
- Dabbs, J. M. Jr. (1982). *Making things visible. Varieties of qualitative research*. SAGE. 31-63.
- Diamond, N., & Graham, H. D. (2000, July/August). How should we rate research universities? *Change*, 10, 21-33.
- Fradd, S. H., & Lee, O. (1997). Teachers' voices in program evaluation and improvement: A case study of a TESOL program. *Teaching and Teacher Education*, 6 (13), 563-577.
- Glastonbury, B., & MacKean, J. (1991) Survey methods. In G. Allan & C. Skinner (Eds.), *Handbook for research students in the social sciences* (225-247). The Falmer Press.
- Hamalainen, K. (2003). Common standards for programme evaluations and accreditation? *European Journal of Education*, 3 (38), 292-300.
- Herman, J. L., Morris, L. L., & Fitz-Gibbon C. T. (1987). *Evaluator's handbook* (2nd Ed.). SAGE Publications, Inc.
- Hughes, R. M. (1925). *A study of graduate schools of America*. The Miami University Press.
- Kayla, C. A., Wheelles, V. E., & Howard, R. D. (1981). *Student evaluation of graduate progress in selected southern universities* (ED 211021). ERIC. Morgantown: Office of Institutional Research, West Virginia University.
- King, J. A., Morris, L. L. , & Fitz-Gibbon C. T. (1987). *How to assess program implementation*. SAGE Publications, Inc.
- Millard, R. M. (1983). The accreditation association: Ensuring the quality of programs and institutions. *Change*, 33-37.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge University Press.
- Patton, M. Q. (1987). *How to use qualitative methods in evaluation*. SAGE.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). SAGE.
- Patton, M. Q. (2003). *Qualitative evaluation checklist*. Retrieved March 18, 2021, from Western Michigan University, Evaluation Center Web site: <http://www.wmich.edu/evalctr/>
- Rowshan, C. (1988). *Student, alumni, and faculty perceptions of quality-related characteristics of the PHD. in English (Rhetoric and Linguistics) at IUP* [Doctoral dissertation, Indiana University of Pennsylvania]. Indiana, PA.
- Scriven, M. (2005). *Key evaluation checklist*. Retrieved March 18, 2021, from Western Michigan University, Evaluation Center Web site: <http://www.wmich.edu/evalctr/>
- Semrow, J. J. (1981). A brief history and background of the accreditation process. *North Central Association Quarterly*, 56, 383-393.
- Sonquist, J. A., & Dunkleberg, W. C. (1977). *Survey and opinion research: Procedures for processing and analysis*. Prentice Hall.
- Stufflebeam, D. L. (2001). Evaluation checklists: Practical tools for guiding and judging evaluations. *American Journal of Evaluation*, 22 (1), 71-79.
- Tezel, K. V. (2006). *A study of an American university master's program in TESOL: Multiple perspectives in program evaluation* [Doctoral dissertation, Middle East Technical University]. Ankara, Turkey. <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- Webster, D. S. (1983). America's highest ranking graduate schools, 1925-1982. *Change* 15, 14-24.
- Worthen, B. R., & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. White Plains.