

Discrepancies in assessing undergraduates' pragmatics learning

Oscar Ndayizeye

Higher Teacher-Training School of Burundi, (Ecole Normale Supérieure (ENS) du Burundi)
Boulevard du 28 Novembre, B.P. 6983 Bujumbura, Burundi
Email: ndaosca@yahoo.fr

Submitted: 15 June 2017 | Revised: 28 December 2017 | Accepted: 03 January 2018

Abstract

The purpose of this research was to reveal the level of implementation of authentic assessment in the pragmatics course at the English Education Department of a university. Discrepancy Evaluation Model (DEM) was used. The instruments were questionnaire, documentation, and observation. The result of the research shows that respectively, the effectiveness of definition, installation, process, and production stages in logits are -0.06, -0.14, 0.45, and 0.02 on its aspect of the assessment methods' effectiveness in uncovering students' ability. Such values indicate that the level of implementation fell respectively into 'very high', 'high', 'low', and 'very low' categories. The students' success rate is in 'very high' category with the average score of 3.22. However, the overall implementation of the authentic assessment fell into a 'low' category with the average score of 0.06. Discrepancies leading to such a low implementation are the unavailability of the assessment scheme, that of scoring rubric, minimal (only 54.54%) diversification of assessment methods, infrequency of the lecturer's feedback on the students' academic achievement, and the non-use of portfolio assessment.

Keywords: *authentic assessment, program evaluation, pragmatics, Rasch model*

How to cite item:

Ndayizeye, O. (2017). Discrepancies in assessing undergraduates' pragmatics learning. *REiD (Research and Evaluation in Education)*, 3(2), 133-143. doi:<http://dx.doi.org/10.21831/reid.v3i2.14487>

Introduction

Writing, for some people, springs out from something else, and the motivation to write this article is remote to 2014 when the authors audited a pragmatics course in English Language and Literature Study Program, Faculty of Languages and Arts of a university. During that time, they observed many but a thing among which the use of (a) classification by (Yule, 1996, pp. 47–48); (b) students' classroom presentations, during which each student was given a sheet used to comment on the presenters' content clarity and the language use in general, and after presentations, students were given a chance to comment/read aloud their reflections on the previous

presentations; (c) a detailed syllabus downloadable from the university's e-data of the staff, giving details on the assessment schema in that course whose assessment comprised students' attendance, class participation, assignments, mid-semester exam (which actually was a take-home exam), and final exam; and (d) a course book written by Yule (1996), entitled *Pragmatics*.

As the authors remarked, the characteristics previously featured are those indicating the authentic assessment of Yusuf (2015, pp. 292–293). However, with this pre-survey insights, Yusuf could not tell whether what he observed was really an authentic assessment being implemented in a pragmatics course. In 2017, wishing to discover more about the au-

thentic assessment as the authors observed that such assessment was quasi-absent in the assessment of linguistics-related course in the first author's country, they decided to go back to the Faculty of Languages and Arts, especially in the 5th semester in which pragmatics course was administered in the English Language and Literature Study Program of the university to investigate the issue.

In (higher) education, the solutions to assessment-related problems can be investigated in a series of aspects, such as, how lecturers may track plagiarism in students' assessment tasks, the development of fair assessment criteria/rubrics, the implementation of authentic assessment, and the impact of students' right to sue educators to the court and how this impedes on assessment. The list of these assessment-related perplexing issues in Indonesian (higher) education system or in the first author's country of origin is far from being exhaustive.

Assessment is a process that is integral part of the logic in which the lecturers' and their students' roles are to be played maximally for the learning to take place. The normal flow is that the lecturers give assessment tasks, and the students do them, and ideally this flow goes on until the students graduate. The problem arises when the two main parties in the teaching-learning process have different perception of some issues.

For example, the views on assessment sometimes diverge as lecturers might view it as a motivation for learning, while their students might see it as the emptiness of any motivation to improve learning but that it is only marking-grounded; and this has also become Fry, Ketteridge, and Marshall's (2009, p. 133) observation. Even among assessors, divergence does also exist. One trend of academics still thrive to use tests (exams) where students give short-answers while another advocates for real-life assignments that result in students' competency, knowledge and interest building. The academics in the last group even label short-answer exams as the traditional practice of assessment. Real-life assignment advocates also stress how this type of activities is related to motivating learning via well-timed and consistent feedback.

Whichever views, it is urgent to see the role of authentic assessment in language classes and how feedback might enhance learning improvement and outcomes in high education. Something obvious is that assessment at this level of education should enhance the students' deep learning approach (Joughin, 2009, p. 19). Getting students to using such approach requires that the assessment tasks be well-prepared.

It should be noted that assessment has attracted and drawn the attention of many academicians and also education practitioners. Some academicians including Mardapi (2008, p. 5, 2012, p. 12) and Fook and Sidhu (2010, p. 153) account assessment as an integral or central part of teaching-learning processes. For instance, Mardapi, in that work, even goes further saying that the efforts to improve the quality of education can be reached through the enhancement of the quality of learning and the quality of its assessment system. The National Research Council [NRC] (1996, p. 5) in DiRanna et al. (2008, p. 8) also insists that assessment and learning are inseparable as they cannot be the two sides of the same coin, which means that the two are mutually inclusive.

The choice of assessment methods has balance some considerations. DiRanna et al. (2008, pp. ix–x) insist that the assessment model should balance and be susceptible (a) to effectively demonstrate how students 'represent knowledge', build knowledge in the course they are learning; (b) to display students' real performance; and (c) to be a good choice of 'an interpretation method' that allows correct inferences about students' performance. If the assessment model choice does not balance the aspects raised above, assessment may not achieve its end in education.

Fry et al. (2009, p. 198) also review how, in the beginning, researching into assessment practices in higher education was not welcome by academicians: they consider such research as either no-need-to-be-done, or as loaded of deliberate disrespect or just one way of treading down their academic space/autonomy. This can be simply considered as 'fearing the unknown' as research can lead to the

extenuating of practices that negatively affect a given educational system as Brown and Glasner (1999, p. 28) stress it. The literature shows that research plays a lot to demonstrate to the academics that they are not geniuses not to need improvements or other new career insights.

The authentic assessment is also related to the notions of the assessor's compliance with assessment principles, formative feedback, scoring rubric, and alignment between learning activities with assessment methods, to quote but a few. It is crucial that some of these key-terms be defined in the context of this research. To begin with, assessment was defined by the University of Queensland, Australia (2007) in Joughin (2009, p. 14) as having to do with any work (which may include assignment, examination, performance or practicum) that is to be completed by a student as a requirement. Assessment is carried for different reasons, ranging from permitting the (1) grading of a student; (2) educational purposes fulfilment, like motivating students' learning, providing necessary feedback to students; and (3) as a student's official achievement record that might be availed as a proof for certification.

The afore-mentioned definition is very clear for it discloses some forms the students' tasks can take, i.e. assessment can be carried out through exams, assignments, practical tasks, and performance. It equally details that assessment has various purposes, i.e. educational and for official record about students' achievement, certifying their competence, and grading them. Educational purpose of assessment will be deepened later. More about the purpose of assessment is proposed by Irons (2008, p. 13). According to him, assessment can serve the purpose of promoting learning through providing helpful feedback, i.e. technically put, through formative assessment and formative feedback.

Feedback, as it appears in the previous line, also needs defining. It is closely related to comments on students' work in order to enhance learning and high learning achievements. According to Irons (2008, p. 13), formative feedback has to do with any piece of information, or simply a process or activity

that is meant to afford or accelerate student learning and this is achieved through comments based on students' outcomes in the formative or summative assessment. The effectiveness of feedback providing depends, among other things, on whether it helps clarify what good performance is (goals, criteria, expected standards) or if it provides opportunities to close the gap between current and desired performance.

It is also important to give account of what authentic assessment is, since the whole study rotates around it. The first view is insisted by Mueller (2014) in Suarta, Hardika, Sanjaya, and Arjana (2015, p. 47) who defines authentic assessment as a form of assessment in which learners demonstrate competence, or a combination of knowledge, skills, and attitude in order to complete an essential task in a real-world situation. Based on this opinion, one can simply put that authentic assessment urges students to make use of their competence or to combine what they have already known with the existent skills just to solve a real-world problem.

Mardapi (2012, pp. 166–167) also accounts for what authentic assessment really is. Madapi stipulates that in this form of assessment, learners present or do a given assignment, the critical thinking is built in the way that students are assessed based on their ability to 'construct' or 'apply' knowledge in a real-world setting, and the evidence of what students are able to do is in live/direct, i.e. it can be observed and this turns authentic assessment to a learner-centered one. The core idea here is that authentic assessment engages students into real-world tasks that incite the use of critical thinking in constructing knowledge.

Another aspect worth underlining is that authentic assessment has got a series of methods that a teacher has to handle given the class size, the students' level of study, and ability. Teachers also smoothly use authentic assessment methods with an aim of aligning teaching-learning activities and tasks, with the assessment method chosen. Diversification of assessment techniques in authentic assessment is demonstrated in the choice offered to teachers. The latter might choose to use stu-

dents' classroom presentations, classroom discussions, individual assignments, group assessments, quizzes, examinations, students' portfolios, students' self-assessment and/or peer-assessment, projects, and performance assessment (Yusuf, 2015, pp. 292–293).

Assessment, especially in high education, is also maximally effective if it complies with a series of principle. In the Indonesian higher education context, the Ministry of Research, Technology, and Higher Education had issued principles as they can be read in the Higher Education Curriculum Book i.e. *Buku Kurikulum di Pendidikan Tinggi* (Tim Kurikulum dan Pembelajaran, 2014, p. 67). According to such a reference, any assessment should be educative, authentic, objective, accountable, and transparent.

In higher education, the literature about the tasks and course objectives alignment, and the assessment methods that enhance learning improvement and outcomes through feedback is still limited. The angle of assessment issue that is still unexploited is how the pragmatics course is assessed authentically given the role was assigned empirically to play for students who will become English language teachers. One among other reasons why only few pragmatics course assessment studies are available is given in McNamara and Roever (2006, p. 54) who comment that assessing a student's ability in pragmatics of a given language is somehow difficult. This is due to the fact that the assessor has to conciliate authentic tasks to be used and practically, given that the necessary costs required to align assessment tasks and practice are huge. However, if some researchers did not explore the angle, this does not mean it cannot be explored.

Rubrics are also great tools to be used in authentic assessment contexts. The rubric formats used in Indonesia, indeed those mentioned in official texts about assessment, are of two types, i.e. descriptive and holistic, and lecturers may choose whichever seems comprehensible to students, efficient and effective in assessing students' knowledge, skills, and competencies. The types and formats of rubrics together with their definitions are available in the *Tim Kurikulum dan Pembelajaran's*

(2014, pp. 69–71) book, in which: (1) rubric is an assessment guide that describes the criteria used by a lecturer in assessing the result of the student's achievement level in his/her assignment/task. In addition, the rubric lists the expected performance characteristics which are manifested/demonstrated in the process and the students' work, and it also becomes a sort of reference to assess each of those performance characteristics; (2) a descriptive rubric provides descriptions of the assessment characteristics or benchmark on each given value scale; (3) holistic rubrics have only one value scale, i.e. the highest scale. The content of the description of the dimensions is the criteria of a performance to the highest scale. If the student does not meet these criteria, the lecturer comments by giving the reasons why the student cannot get the maximum score in his/her tasks.

It should be noted that the low quality of rubrics, indeed any rubric which is not clear, or simply wrongly constructed climaxes in doubts about the scoring integrity of the assessor concerned. Further, Christie et al. (2015, p. 31) investigate how assuring assessment grading tools quality affects student motivation and learning. The study displays how the Australian and USA lecturer's assessment practices of not using scoring rubrics to assess the quality of students' work tend to turn the final judgment of students' learning into a questionable one. The lecturers involved in that study tend to use common sense in assessment scoring instead of written rubrics, which could affect negatively, as the authors observed, the lecturer's integrity in grading students' work. With such conviction in mind, this study investigated the still-unexploited angle of assessment issues, that is, how pragmatics course is assessed authentically given its importance for the teacher students of English language. This research was sorely concerned with the implementation of authentic assessment in higher education. Some related aspects such as alignment, feedback, and compliance with the assessment principles are also tackled.

The problem was formulated around the idea of curiosity to know the extent to which the authentic assessment was imple-

mented in the pragmatics course taken by semester five students in the English Language and Literature Study Program. Since such an assessment has its own indicators, the problem also includes: (1) how the assessment standard is indicated in the curriculum being implemented in the pragmatics course, (2) the proof of alignment between students' tasks and the assessment methods in the pragmatics course, (3) the pragmatics course assessment methods providing more feedback to the students, (4) what the compliance with the authentic assessment principles in assessing students' tasks in the pragmatics course is like, and (5) what the authentic assessment implementation in the pragmatics course is like.

Carrying out this program evaluation was beneficial, *firstly*, to the theoretical literature by broadening it as far as the evaluation of the implementation of the authentic assessment in teaching pragmatics course to Indonesian students who are expected to be teachers of English language is concerned. Equally, this work is meant to broaden more literature regarding the use of the Discrepancy Model of Evaluation (DME) in foreign language assessment, especially in English as a Foreign Language (EFL) settings. *Secondly*, it is also beneficial to the practical aspect, because the students who are taking the pragmatics course might foster some new ideas to the pragmatics course lecturer in the perspective of adjustment as far as the course administration is concerned. Furthermore, broader space is also open to other researchers to investigate into the realms of authentic activities and assessment that might develop EFL teacher students' pragmatic competence, especially the pragma-linguistic and also socio-pragmatic competencies.

The research questions in this study were based on the problem formulated and the DEM stages, i.e. pragmatics course Program Definition, Installation, Process, and also Product (Fernandes, 1984; Fitzpatrick, Sanders, & Worthen, 2011, pp. 156–157). Those questions are: (a) to which degree did the assessment that was carried out in the pragmatics course comply with the authentic assessment standard as indicated in the curriculum? (b) what is the proof of alignment be-

tween the assessment methods used in the pragmatics course and the students' learning activities? (c) what were the most consistent feedback providing assessment methods among the ones used in the pragmatics course assessment? (d) what were the possible necessary inputs for the implementation of the authentic assessment carried out in the pragmatics course? (e) to which extent had the authentic assessment been implemented in the pragmatics course?.

Method

This research is a program evaluation that employed Provus's Discrepancy Evaluation Model. This program evaluation was carried out at a university which is located in Yogyakarta Special Region, Indonesia. The population of this study was the semester 5 pragmatics course takers. The research employed non-probability sampling method and saturated sampling technique (in which population is equal to sample) was used with $n=31$.

Procedure

The core is that there is a determination of: (1) the Standard (S), i.e. how the pragmatics course assessment should be conducted, based on the Ministry of Research, Technology, and Higher Education assessment principles as stated in the Higher Education Curriculum Book (*Tim Kurikulum dan Pembelajaran*, 2014, pp. 67–74), i.e. *Buku Kurikulum Pendidikan Tinggi* and the university's English Language and Literature Study Program Curriculum (2014), and then (2) taking Performance (P) measure, i.e. given the pragmatics course inputs/resources, at this stage, the pragmatics course assessment characteristics were observed, and the assessment process was scrutinised. Then, it was followed by the evaluation *per se*, i.e. the determination of discrepancies (D) by comparing Performance (P), i.e. how the program performs compared to the Standard (how it should behave).

Data, Instruments, and Data Collecting Technique

In the pragmatics course program evaluation, both quantitative and qualitative data

were collected. Three instruments were used in order to collect the data in this study, including: questionnaire, observation guide, and documentation. Through the questionnaire, the data about the assessment techniques, most feedback providing technique, compliance with assessment principles, resources, and the effectiveness of each assessment technique in uncovering the students' ability were collected. By documentation, information about the pragmatics course objectives, assessment standards, the rubrics used, and students' final learning outcomes were gathered. The observation instrument helped the authors in gathering information about the main inputs (curriculum, lecturer, and students), the assessment methods used, details about the assessment process, and teaching-learning facilities.

Data Analysis Techniques

Two types of analysis were carried out, i.e. (descriptive) quantitative analysis through Rasch Model with the Winsteps software version 3.73.0 and qualitative analysis: following Miles, Huberman, and Saldana (2014, pp. 12–13) technique consisting of (1) data reduction or condensation, (2) data display, and (3) conclusion drawing/verification.

Evaluation Criteria

Table 1 shows the the criteria of the level of authentic assessment implementation.

Table 1. (Dis)agreement and authentic assessment level of implementation

Interval	Categories
$X < -0.99$	Strongly Agree/Very High
$-0.99 \leq X \leq 0$	Agree/High
$0.1 \leq X \leq 1.01$	Disagree/Low
$X \geq 1.01$	Strongly Disagree/Very Low

(Developed based on Sumintono and Widhiarso (2015, p. 40)

Note:

X: stands for each statement's 'Item Measure' value in logits as analysed through Winsteps version 3.73.0.

Meanwhile, Table 2 provides the information about students' scores categorization.

Table 2. Categorizing the students' scores

Score X	Categories	Criteria
$X \geq M + 1. SD$	Very High	$X \geq 3$
$M \leq X < M + 1. SD$	High	$2.5 \leq X < 3$
$M - 1. SD \leq X < M$	Low	$2 \leq X < 2.5$
$X < M - 1. SD$	Very Low	$X < 2$

Source: Mardapi (2008, p. 123)

Note:

M : Mean of students' final scores in the pragmatics course

X : Each single student's score out 4 (because the score scale is 4-1)

SD : Standard deviation; obtained through $SD = (4-1)/6$ as the score scale is 4-1

In order to admit that a given method was used, it has to satisfy the criteria that: Mean=1 (or close to 1, that is 0.9), and $STD \leq 0.31$. Similarly, to determine whether there has been diversification of assessment methods and the students' success rate in the pragmatics course, some criteria were used:

<50%	: Low
50%-65%	: Average/Minimal
66%-81%	: High
$\geq 82\%$: Very High

Findings and Discussion

Before the results and discussion is presented, it should be underlined that item measure values for quantitative data are expressed in logits. For Rasch model applied in social sciences, the more the item measure value in logit gets superior to 0, the more the subjects do not agree with the statements presented to them. On the contrary, if the item measure value is equal to 0 or negative, this is an indication that the statement was agreed on by the respondents. In few words, the logit values comprised between -2 up to ≤ 0 are indicators that statements concerned are agreed by the respondents.

The discussion starts with quantitative data followed by qualitative data. Concerning the quantitative data, at the program Definition Stage, the resources/inputs recognized by the pragmatics course takers as primordial included: the lecturer, course objectives, classroom ability to cater for all the students, class

cleanness, sufficiency of chairs, adjustable luminosity, functional fans, and also LCD projector as their measure values in logits are respectively -0.79, -0.26, -0.57, -0.16, -0.79, -1.00, -1.00, and -1.23.

At the pragmatics course Installation Stage, the following is the comparison between the standard performance of the program and how it should behave. It is an activity aimed at finding the discrepancies. Given the pragmatics Program Process Stage/Assessment process, the performance of the program has indicators of good performance in terms of the assessment principles of being educative, authentic, and the alignment of learning activities with the assessment used. Based on the measure values related to the positive indicators of good performance, the following measure values are more illustrative: -0.26, -0.16, -0.16, -0.16, and -0.57. It should be noted that the values represent respectively the fact that the assessment principles of being educative and authentic, the last three values are concerned with the statements about alignment.

The latter was accepted as having been observed by the lecturer of pragmatics. By doing so, she complied with the guideline which was provided in the study program curriculum, Higher education (HE) (Tim Kurikulum dan Pembelajaran, 2014), citing the Ministry of Education and Culture's Decree Number 49 of 2014, article 20, about HE in Indonesia, Sections 1 and 4 about assessment in HE.

Nevertheless, the core activity at DEM program of Installation is finding discrepancies, those which have been registered are non-compliance with the assessment principles of objectivity, accountability, and implicitly that of feedback. The item measure values associated with those three principles are superior to 0.1. The score fits to the criterion of $0.1 \leq X \leq 1.01$, so that it indicates that the respondents disagreed that there was optimization of the three principles previously mentioned. There was no use of portfolio assessment although it was recommended in the English Language and Literature Study Program and High Education Curriculum Book (*Buku Kurikulum di Perguruan Tinggi*). As portfolio is described as a highly-recommended

assessment method that allows lecturers to keep an eye on every student's knowledge process in the study program curriculum, if this lack is added to infrequency of feedback by the lecturer, the fact of not using portfolio was felt as a discrepancy.

The DEM program process stage is concerned with the results of the mostly used authentic assessment methods, the extent to which assessment methods were diversified, and the authentic assessment method, one of which is was the most feedback providing. On the list of the eleven authentic assessment methods found in the literature, six were admitted to have been used in the pragmatics course. The criteria used in determining that a given assessment method was used are that of Mean = 1, and $SD \leq 0.31$. The following authentic assessment methods are satisfying: students' classroom discussion, individual assignments, quizzes, examinations, project assessment, and group assignments. The descriptive statistics (mean; SD) features are respectively: (1;0), (1;0), (0.90; 0.31), (1;0), (1;0), and (1;0). If these values are compared to the criteria pre-established, the aforementioned authentic assessment methods satisfied them thoroughly.

The second aspect looked at this point was authentic assessment method diversification. Simple calculations showed that the diversification was but average/minimal. Over the total of eleven authentic assessment methods, if six only were used, this means that the diversification was of $(6 \times 100) / 11 = 54.54\%$. Compared to the criteria, this percentage falls into the 50%-65% interval, which is signifying that such diversification is simply 'Average/Minimal'.

On the top of that, the respondents' appreciation of group assignment assessments is shown in two ways: (1) they agree that it provides them with valuable feedback; (2) they recommend it to the lecturer for a better administration of pragmatics course in the future. This is indicated by its related item measure value in logits, which is -0.47. If such measure is compared to the criteria set, this illustrates that group assignments were admitted to have provided helpful feedback to the pragmatics course takers. Such finding is in

line with Bentley and Warwick (2013). Lately, students appreciate group assignment assessment as they gain learning from their friends/peers and develop teamwork, communication, and also interpersonal skills.

Furthermore, the respondents recommend the use of group assignments, one of the techniques of authentic assessment, to the pragmatics course lecturer. This is also a case in Fook and Sidhu's (2010) study that sought to examine the implementation of authentic assessment in higher education in Malaysia, especially in the course of 'Testing, Assessment, and Evaluation 752' (TSL 752) which is taught in a Master Program at the Faculty of Education of a public university in Selangor, Malaysia. In both of these studies, authentic assessment was proven as being susceptible or appreciated to enhance learning as it won acceptance from the respondents.

Students who are successful in the pragmatics course have the scores ranging from 2.5 to 4 as it is well-described in the students' academic guide which is termed *Peraturan Akademik* (Universitas Negeri Yogyakarta, 2014, p. 15). Except for two students who were in irregular conditions, 29 out of 31 students got a score comprised between 2.66 and 4. Compared to the criteria pre-set in Table 1, students' scores fall in 'High' and 'Very High' categories.

As far as the qualitative data are concerned, the analysis led to the observation that the pragmatics course lacked clear assessment and scoring scheme, and the fact of not using portfolio although it is described as a highly-recommended assessment method that allows the lecturers to keep an eye on every student's knowledge process. The infrequency of the lecturer's feedback to students' learning and assignments was also found.

Similar findings were found in Christie et al. (2015, p. 31). Later, it is demonstrated that Australian and USA lecturer's assessment practices of not using scoring rubrics to assess the quality of the learners' work tend to turn the final judgment of students' learning into a questionable one. Simply put, if the respondents'/students' perceptions are that there was no maximization of the objectivity and accountability principles in that course,

the students might have suspected the scoring integrity.

In general, the evaluation result of each stage is presented in Table 3.

Table 3. Holistic evaluation of authentic assessment implementation

No	DEM Stage/ Component	Average	Category
1	Program Definition Stage	-0.06	High
2	Program Installation Stage	-0.14	High
3	Program Process Stage	0.45	Low
4	Program Product Stage	0.02	Low
Average for the 4 DEM Stages		0.06	Low

The students' final scores in the pragmatics course are averaged and categorized as follows:

Average : 3.22

Category : Very High

Therefore, the pragmatics course definition and product (based on the students' scores aspect) are respectively in 'High' and 'Very High' categories as the average for the item measure order value for the DEM Definition stage is -0.06, while the average for the students' final scores is 3.22. The performance of the pragmatics course over the resources/inputs is also in 'High' category. Such performance is not maximal as explained by the DEM Process Stage which has the average for the item measure order value of 0.45, falling then in 'Low' category. Another aspect of the DEM product stage (concerned with the effectiveness of assessment methods used in uncovering the students' knowledge, ability, and competence) is in 'Low' category with the average for the item measure order value of 0.02.

Conclusion and Suggestions

Conclusion

A general overview of the implementation of authentic assessment is in 'Low' category. The definition and installation stages

are in 'High' category. One aspect of the pragmatics course product stage is in 'Low' category because the process itself is stained by some impediments and it is in 'Low' category. The diversification of the assessment methods is still 'Average/Minimal'. That conclusion is formulated by the following main findings. Firstly, the compliance of the pragmatics course assessment with the curriculum assessment standard is found to be in 'High' category. However, at the DEM Pragmatics Installation Stage, the discrepancies registered: (a) are little compliance with the assessment principles of feedback, objectivity, and also accountability; (b) lack the pragmatics assessment plan and scoring rubrics; (c) lack tasks and assessment methods that will push students for further research in the field of pragmatics; (d) are ineffective to support students' learning monitoring due to no use of portfolio assessment. Secondly, the proof of alignment of students learning activities and assessment methods is that: (a) the students' intended learning outcomes are in line with the study program curriculum; (b) the problem-solving skills which are engaged by the students during the learning activities resemble those required to solve assessment tasks. Thirdly, the most consistent feedback providing assessment method is group assignments. Meanwhile, the other assessment methods which are used include: (a) students' classroom discussion, (b) individual assignments, (c) quizzes, (d) examinations and also project assessment. Fourthly, the inputs which are found to be necessary for the implementation of the authentic assessment in the pragmatics course to be possible course include: (a) the lecturer, (b) the course objectives, (c) the classroom that is clean and big enough to cater for all the students, (d) enough chairs, (e) adjustable luminosity, and also (f) functional fans and LCD projector. Fifthly, the level of implementation of the pragmatics course is transcribed in the DEM Pragmatics Course Product stage that includes two aspects of the product: (a) effectiveness of the assessment methods in uncovering the students' ability, which is in 'Low' category, (b) the students' final scores in the pragmatics course, which are in 'Very High' category.

Implications

Based on the conclusions, the implications for practice are: (1) until the teachers/lecturers choose activities that push students to use available learning resources, students will always perceive such expensive resources or services as having less importance in their learning; (2) until used up teaching/learning resources are replaced, they are seen as inexistent by students; (3) the lecturer's teaching effort and high academic competence without availing a clear assessment scheme and a scoring rubric might stain the whole scoring integrity for that teacher; (4) lecturers may use many assessment methods, and there may be alignment between students' learning activities and expected outcome assessment methods, but still assessment methods providing valuable feedback to students being very few; (5) a course where students' success rate is high as indicated by students' final scores does not implicate that the whole assessment practice has been without any spot mark.

Suggestions

Suggestions for the university administration, lecturers, and educational researchers or education practitioners are as follows. (1) The university's administration should conduct a regular check of the used-up learning resources in the classroom and replacement of those in bad conditions. (2) The pragmatics course lecturers are suggested to (a) apply the more student-centred teaching approach (more interactive and more chance for students to talk); (b) choose students' learning activities that push them to learn how to use resources provided by the university. (It would be unfortunate that the university presumably pays much for external journals and the Internet hotspot maintenance, but the students still say that those resources do not improve their pragmatics course learning); and (c) explain and give students opportunities to ask about either the tentative or provisional assessment scheme as well as scoring rubric. (3) Other researchers are suggested to (a) carry out other studies to evaluate the implementation of authentic assessment in the English Language and Literature Study Program particularly and all the FLA (Foreign Language Assistant) de-

partments generally, and (b) conduct other research related to the lecturers' teaching strategies/techniques, methods, and learning activities. (4) There should be a development of a model of applying Item Response Theory or any model linked to it (e.g. Rasch model) in the assessment practices in Indonesian higher education.

References

- Bentley, Y., & Warwick, S. (2013). An investigation into students' perceptions of group assignments. *Journal of Pedagogic Development*, 3(3), 11–19. Retrieved from <https://journals.beds.ac.uk/ojs/index.php/jpd/article/view/199/310>
- Brown, S. A., & Glasner, A. (1999). *Assessment matters in higher education: Choosing and using diverse approaches*. Buckingham: Society for Research into Higher Education & Open University Press.
- Christie, M. F., Grainger, P., Dahlgren, R., Call, K., Heck, D., & Simon, S. (2015). Improving the quality of assessment grading tools in Master of Education courses: A comparative case study in the Scholarship of Teaching and Learning. *Journal of the Scholarship of Teaching and Learning*, 15(5), 22–35. <https://doi.org/10.14434/josotl.v15i5.13783>
- DiRanna, K., Osmundson, E., Topps, J., Barakos, L., Gearhart, M., Cerwin, K., ... Strang, C. (2008). *Assessment-centered teaching: A reflective practice*. Thousand Oaks, CA: Corwin Press.
- Fernandes, H. J. X. (1984). *Evaluation of educational programs*. Jakarta: National Education Planning Evaluation and Curriculum Development.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines*. Boston, MA: Pearson Education.
- Fook, C. Y., & Sidhu, G. K. (2010). Authentic assessment and pedagogical strategies in higher education. *Journal of Social Sciences*, 6(2), 153–161. <https://doi.org/10.3844/jssp.2010.153.161>
- Fry, H., Ketteridge, S., & Marshall, S. (2009). *A handbook for teaching and learning in higher education: Enhancing academic practice* (3rd ed.). New York, NY: Routledge.
- Irons, A. (2008). *Enhancing learning through formative assessment and feedback*. London: Routledge.
- Joughin, G. (Ed.). (2009). *Assessment, learning and judgement in higher education: A critical review*. Wollongong: Springer. https://doi.org/10.1007/978-1-4020-8905-3_2
- Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendekia.
- Mardapi, D. (2012). *Pengukuran penilaian dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- McNamara, T. F., & Roever, C. (2006). *Language testing: The social dimension*. Oxford: Blackwell Publishing.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.
- Suarta, I. M., Hardika, N. S., Sanjaya, I. G. N., & Arjana, I. W. B. (2015). Model authentic self-assessment dalam pengembangan employability skills mahasiswa pendidikan tinggi vokasi. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 19(1), 46–57. <https://doi.org/10.21831/pep.v19i1.4555>
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada asesmen pendidikan*. Cimahi: Trim Komunikata.
- Tim Kurikulum dan Pembelajaran. (2014). *Buku kurikulum pendidikan tinggi*. Jakarta: Directorate of Learning and Student Affairs, Directorate General of Higher Education, Ministry of Education and Culture.
- Universitas Negeri Yogyakarta. (2014). *Buku peraturan akademik Universitas Negeri*

- Yogyakarta (Revised ed.). Yogyakarta: UNY Press.
- Yule, G. (1996). *Pragmatics*. Oxford: Oxford University Press.
- Yusuf, A. M. (2015). *Asesmen dan evaluasi pendidikan: Pilar penyedia informasi dan kegiatan pengendalian mutu pendidikan*. Jakarta: Prenada Media Group.