A Comparative Analysis Between Good Feedback Descriptors on Online Courses

Anderson Pinheiro Cavalcanti[1,3]**, Vitor Belarmino Rolim**[3]**,**
Dragan Gašević[4]**, Rafael Ferreira Mello**[2,4]

[1]SiDi – Recife – PE – Brazil

[2]Cesar School – Recife – PE – Brazil

[3]Centro de Informática – Universidade Federal de Pernambuco
Recife – PE – Brazil

[4]Faculty of Information Technology, Monash University,
Melbourne, Australia

`{apc,vbr}@cin.ufpe.br, rflm@cesar.org.br`

`dragan.gasevic@monash.edu`

Abstract. *Feedback is a critical component of the teaching-learning process. Through it, teachers share relevant information so that students understand the subjects and activities, in addition to promoting self-regulation. However, the activity of writing and sharing feedback is not easy and may even lead to students' demotivation. Given this, it is possible to find several works in the educational research literature that propose good practices for elaborating textual feedback. This work aims to analyze the relationship between different models of good practices for feedback and the relationship between English feedback and Portuguese feedback using an epistemic network (ENA). The results show that some characteristics are similar, but the models do not have a direct relationship. It is concluded that a combination of these models can improve the analysis of feedback quality.*

1. Introduction

Feedback is a mechanism that plays a critical role in the instructor-student relationship. This mechanism allows students to improve their skills and competencies and allows instructors to tailor their methods and content based on students' learning needs. Moreover, instructors use the feedback messages to suggest insights to improve students' performance by enhancing interaction, assessments and answering questions [Langer 2011, Matcha et al. 2019]. According to [Sadler 1989], feedback needs to provide precise information about the learning task or process that fills a gap between what is understood and what should be understood. Several studies have shown that useful feedback brings benefits to learning [Hattie and Timperley 2007, Nicol and Macfarlane-Dick 2006, Parikh et al. 2001].

[Hattie and Gan 2011] highlighted that feedback is one of the top ten aspects of learning to enhance student performance and learning experience. According to the study by [Sadler 2010], feedback only works when applied to fill the gap between the current and desired performance. If the information shared by feedback is not or cannot be processed by the student to produce improvements, it will not affect learning. Furthermore,

inadequate feedback can negatively impact students, creating distrust in the feedback process and in the teaching, which can impair students' self-efficacy and motivation and even result in dropout [Boud and Falchikov 2007].

Some works in the literature present feedback descriptors for the instructor to provide good feedback. [Nicol and Macfarlane-Dick 2006] propose a self-regulation model and identify seven principles of good feedback practices. Each principle has specific strategies the teacher can use to facilitate student self-regulation. On the other hand, [Hattie and Timperley 2007] proposed a feedback model to identify the circumstances in which feedback has the most significant impact. This model has four levels where feedback can be provided (task, processing task, self-regulation, and personal). The key idea of this work is to ensure that feedback is provided at the appropriate level, making it effective feedback.

Recently, several studies have proposed machine learning techniques designed to identify the good practices automatically [Nicol and Macfarlane-Dick 2006] and levels [Hattie and Timperley 2007] from feedback shared through educational technology. In general, these techniques adopted random forest classifiers in combination with features related to text structure to categorize feedback messages according to the good practices and levels of feedback [Cavalcanti et al. 2021, Osakwe et al. 2021]. Although these studies indicate a promising direction to create an automatic tool to support instructors in creating better quality feedback messages, they do not analyze the relationship between the good practices [Nicol and Macfarlane-Dick 2006] and levels [Hattie and Timperley 2007]. Analyzing relationships between these two models is essential as it can inform the assessment of feedback quality; that is, whether it is sufficient to use only one of these two well-known feedback models or their combined use is preferred.

Therefore, this paper proposes a study that adopts epistemic network analysis (ENA) to perform a comparative analysis between the feedback descriptors proposed by [Nicol and Macfarlane-Dick 2006] and [Hattie and Timperley 2007]. ENA provides techniques for quantitative analysis of qualitative aspects of learning and teaching [Shaffer 2017], as it has been shown through the growing adoption of ENA in learning sciences, learning analytics, and educational research [Shaffer et al. 2009, Rolim et al. 2019]. With this analysis, we intend to produce new theoretical insights into providing good quality feedback in technology-enhanced learning.

2. Background

Previous works have proposed models or principles that help increase the impact of feedback on student learning. For example, [Hattie and Timperley 2007] proposed a model to assist the construction of effective feedback. This model identifies three main questions that effective feedback must answer: "*Where am I going?*", "*How am I going?*", "*Where to go next?*". [Evans 2013] performs an extensive literature review and proposes a feedback scenario that identifies specific areas for future research on evaluation feedback in higher education. [Nicol and Macfarlane-Dick 2006] presents seven principles of good feedback practices to assist teaching staff. The works of [Hattie and Timperley 2007] and [Nicol and Macfarlane-Dick 2006] have well-formulated models, are widely used in the literature, and are detailed below.

2.1. Seven Good Feedback Practices

[Nicol and Macfarlane-Dick 2006] proposed a conceptual model of self-regulation based on a review of the research literature on formative assessment and feedback. The key idea of the work is to identify how formative assessment and feedback processes can help promote self-regulation. Based on the conceptual model, seven principles of good feedback practices were defined that the instructor can use to think about the project and evaluate his feedback procedures. The seven good feedback practices are [Nicol and Macfarlane-Dick 2006]:

- **GP 1**: Helps clarify what good performance is (goals, criteria, expected standards);
- **GP 2**: Facilitates the development of self-assessment (reflection) in learning;
- **GP 3**: Delivers high-quality information to students about their learning;
- **GP 4**: Encourages teacher and peer dialogue around learning;
- **GP 5**: Encourages positive motivational beliefs and self-esteem;
- **GP 6**: Provides opportunities to close the gap between current and desired performance;
- **GP 7**: Provides information to teachers that can be used to help shape teaching;

Each feedback principle could be connected to different strategies that the instructor can implement in the classroom. According to [Nicol and Macfarlane-Dick 2006], good feedback practices are broadly defined as any strategy or content that can increase students' ability to self-regulate their learning performance.

2.2. Four Feedback Levels

[Hattie and Timperley 2007] proposed a feedback model that identifies the circumstances where feedback has a higher impact. The authors define the purpose of feedback as reducing discrepancies between current performance and the desired objective. The authors looked at several conditions that could maximize the positive effects of feedback on learning, including increasing student awareness of an overall learning goal, progress towards the goal, and the subsequent goals necessary to achieve the primary goal. The proposed model specifies three types of questions that feedback is designed to answer (Where do I go? How do I go? Where next?). Moreover, each feedback question operates at four levels: task feedback (FT), feedback on the task processing (FP), feedback on self-regulation (FR) and feedback on the self as a person (FS). The goal of each level is shown below.

- **FT Level**: Feedback can be about a task, such as whether the job is correct or incorrect, can include instructions for more or different information.
- **FP Level**: Feedback can be directed to the process used to create a product or complete a task, is more directed to information processing or learning processes that require understanding or completing the task.
- **FR Level**: Feedback for students can be focused on the level of self-regulation, including greater self-assessment or confidence skills, which can have major influences on self-efficacy, self-regulatory proficiency, and students' personal beliefs as learners.
- **FS Level**: Feedback can be personal in the sense that it is directed to the self. It is often unrelated to task performance.

The authors state that feedback on the FS level is the least effective. The FR and FP levels are powerful in terms of deep processing and task domain, while FT is powerful when feedback is used to improve strategy processing or improve self-regulation [Hattie and Timperley 2007], that is FT is powerful when used in combination with either FP and FR.

2.3. Related works

Several previous papers in the literature focused on the effectiveness of feedback in educational environments. According to Gibbs and Simpson [Gibbs and Simpson 2005], formative assessment combined with high-quality feedback can have a powerful impact on student learning. The following are some related works that guided this study.

The work of Dunworth and Sanchez [Dunworth and Sanchez 2016] presents the results of an embedded multiple case study that investigated teaching staff and students' views on written feedback at a United Kingdom university. The study results indicated a general consistency between the teaching staff and students regarding the nature of quality feedback. Therefore, Dunworth and Sanchez state that quality feedback can be described as a process in which teaching's contribution and support are actively engaged, appropriated, and used productively by students to improve their educational experience from several different dimensions (affective or interpersonal, orientational and transformational). The results suggest that feedback needs to incorporate each of these dimensions to be perceived as good quality.

Epistemic Network Analysis (ENA) has been used in the literature to analyze and visualize the relationship between data [Shaffer et al. 2016]. This technique provides tools to analyze data qualitatively and has been used in the educational field; for example, [Rolim et al. 2019] presents an approach that uses ENA to understand the relationships between cognitive and social presence. The paper demonstrates how epistemic network analysis (ENA) can provide new qualitative and quantitative insights into the development of students' social and critical thinking skills in communities of inquiry. In addition, the authors explore how the relationship between social and cognitive presences changed over time during a course.

[Mello and Gašević 2019] investigated different configurations of ENA parameters to analyze student interaction behavior in online discussions from the perspective of the research community framework. In their results, the authors state that the main implication was that removing dominant codes can lead to a better understanding of one of the two key constructs in the Community of Inquiry model (social presence), but reduce the perception of the other (cognitive presence).

[Ferreira et al. 2018] analyzes students' cognitive development in asynchronous discussions in online learning environments. The authors combine natural language processing techniques and graph-based analysis with ENA to provide qualitative insight into developing students' critical and deep thinking skills. [Gašević et al. 2019] propose an approach called SENS (epistemic social network signature). This approach aims to combine social network analysis (SNA) with epistemic network analysis (ENA) to model the different dimensions of collaborative learning, using social ties and content analysis of student messages.

In this context, this paper proposes a network analysis method to uncover the

relationship between different constraints of good quality feedback. In this paper, we evaluate the relationship between the concepts proposed by [Hattie and Timperley 2007] and [Nicol and Macfarlane-Dick 2006], but the same method could be generalized to any other theory.

3. Research Question

The previous sections demonstrate that feedback is critical to help self-regulate student learning. The feedback models proposed in [Hattie and Timperley 2007, Nicol and Macfarlane-Dick 2006] have been widely used in the literature. However, the relationship between the constraints proposed by both models was not explored. Therefore, our research question that guides this study is:

> **RESEARCH QUESTION (RQ):** *What are the relationships between indicators of the feedback models proposed by [Hattie and Timperley 2007] and [Nicol and Macfarlane-Dick 2006]?*

With this research question, we posit that the analysis of feedback quality can be far more insightful if feedback quality is assessed from two different and complementary perspectives.

4. Method

4.1. Datasets

This work used two datasets, with feedback in Portuguese and English. Each one of them is detailed below.

4.1.1. Portuguese Dataset

The Portuguese dataset contains individual written feedback provided by instructors who posted the feedback messages through the activity submission tool in the LMS. The dataset comprises 1,000 examples of feedback from 2 courses: biology (41 examples of feedback) and literature (961 examples of feedback). The average length of feedback texts is 30 words. Each feedback was classified by expert coders according to the 11 feedback descriptors (7 good practices [Nicol and Macfarlane-Dick 2006] and four levels [Hattie and Timperley 2007]). The feedback message received labeled 0 if it did not belong to the descriptor or label 1 if it belonged to the descriptor. This annotation process was done separately between the seven good practices and the four levels.

The measures used to analyze the quality of the annotation process were the percentage of agreement and Cohen's kappa coefficient (κ). According to [Landis and Koch 1977], Cohen's kappa (κ) is a statistic that measures inter-rater agreement for qualitative items, where values close to 1 indicate significant agreement between the raters or coders (in this case, expert classifiers) and values close to 0 indicate that the agreement between the raters or coders is purely random. The inter-rater agreement was moderate for the seven good practices, reaching 76.7% and Cohen's $\kappa = 0.43$. For the four levels, the inter-rater agreement had a percentage of 72.2%, and Cohen's κ was 0.44. Another two experts who did not participate in the first coding stage resolved the divergent cases, 23.3% for the seven good practices and 27.8% for the four levels. Table 1 shows the Portuguese dataset division for the seven good practices and the four levels.

Tabela 1. Portuguese dataset division for the seven good practices and the four levels.

	Good Practice (GP)							Level			
	GP1	GP2	GP3	GP4	GP5	GP6	GP7	FT	FP	FR	FS
class 0	875	971	692	987	766	397	1000	112	499	992	849
class 1	125	29	308	13	234	603	0	888	501	8	151
Total	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

4.1.2. English Dataset

The English dataset has 272 feedback texts from instructors of two courses: Software Engineering and Learning Analytics. The average length of feedback texts is 437 words. The annotation process followed the same process as the Portuguese dataset. For the seven good practices, the inter-rater agreement had a percentage of 76.6%, and Cohen's κ was 0.53. For the four levels, the inter-rater agreement had a percentage of 63.8% and Cohen's κ was 0.38. Different codders who did not participate in the first annotation step resolved the divergent cases. Table 2 shows the English dataset division for the seven good practices and the four levels.

Tabela 2. English dataset division for the seven good practices and the four levels.

	Good Practice (GP)							Level			
	GP1	GP2	GP3	GP4	GP5	GP6	GP7	FT	FP	FR	FS
class 0	193	206	38	166	80	202	253	94	68	174	139
class 1	79	66	234	106	192	70	19	178	204	98	133
Total	272	272	272	272	272	272	272	272	272	272	272

4.2. Network Analysis

To answer our research question, we use Epistemic Network Analysis (ENA) [Shaffer et al. 2009] to analyze the relationship between feedback levels and good practices of feedback. ENA is a graph-based analysis technique to model patterns of association by a network of dynamic relationships of different concepts [Shaffer et al. 2016]. Within ENA, the network of relationships among different *codes* is created for each *unit of analysis*. Two codes are considered related if they appear in the same chunk of text, called *stanza* (or *conversation*). We can see ENA relationships between the two feedback models, i.e., if the good practices units are near and strongly connected with the feedback levels units.

In this work, the *units of analysis* are feedback messages, the *codes* were the levels of feedback and the good practices of feedback, whereas the *stanza* we considered individual feedback messages. The network data provided by ENA was visualized as undirected graphs, and the object of interest was the interactions between the elements [Shaffer et al. 2016]. In this network, there were three elements to observe: the *size* of the nodes represents their frequency; the *nearness* of the node represents the similarity among them; the *strength* of the code relationship represents the frequency of their co-occurrence.

5. Results

To answer our research question, we analyzed the ENA network using: Portuguese dataset (Figure 1), English dataset (Figure 2), and both datasets (Figure 3), respectively.
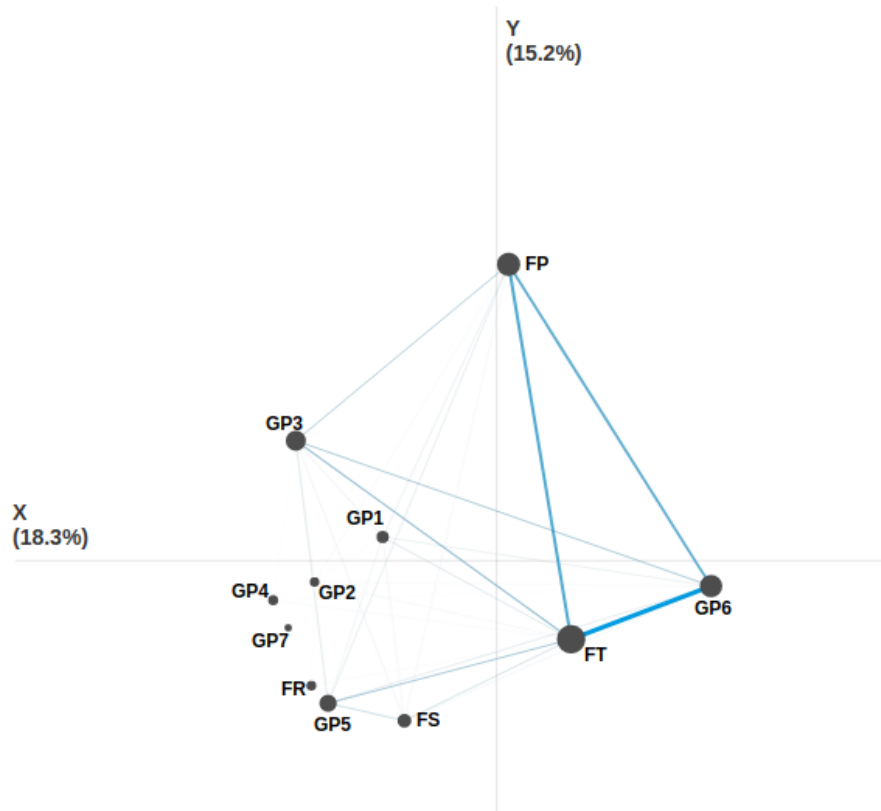

Figura 1. ENA network of the relationship between the feedback descriptors for Portuguese.

Figure 1 presents the ENA graph for the Portuguese language. It shows that the most substantial relationship in the graph is between the FT and GP6 classes. The GP6 good practice aims to provide opportunities to close the gap between current and desired performance, where the instructor should indicate to the students their mistakes to help them avoid the same mistake in future tasks. Similarly, the FT level aims to provide comments that indicate where the student got it right/wrong on a specific task. In other words, feedback of these classes has a common goal: to show the student the points where he hit or miss the activity and, consequently, help the student close the gap between current and desired performance.

Another strong relationship in the graph is between FP and FT and between FP and GP6. An interpretation is that the FT and FP levels are related to a specific activity. The difference between these levels is that the FT focuses only on the task, checking whether it is correct or incorrect, and the FP focuses on the processing of the task (i.e., the student must reflect on his answer). Similar to our justification of the connection between FT and GP6, the connection between FP and GP6 shows that feedback about the task makes students reflect and also helps close the gap between current and desired performance.

The GP3 good practice that aims to provide students with high-quality information

had a relationship with the FT and FP levels. As the FT and FP levels provide feedback on the tasks and processing, this means that on these levels, high-quality information is provided, and it helps the student to identify not only where the student got it wrong or right but why they got it wrong/right and what are the next steps for learning.

It is also possible to see in the graph, through proximity and connection, the relationship between GP5 and FS. The goal of GP5 good practice is to encourage positive motivational beliefs and self-esteem. This relationship explains the relationship with the FS level because, on this level, feedback must provide personal praise for the student to remain motivated in his learning.
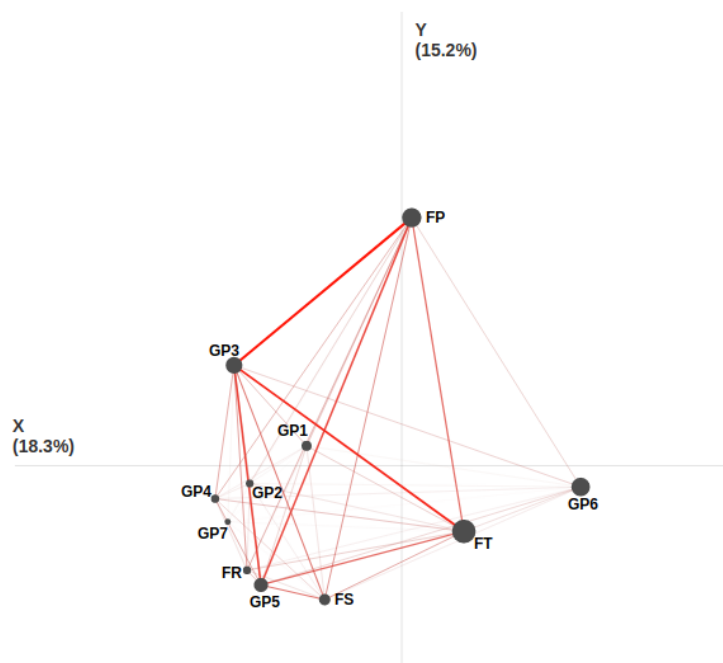

Figura 2. ENA network of the relationship between the feedback descriptors for English.

Figure 2 shows the ENA graph for feedback messages written in English. It is possible to see that the connections are well distributed, differently from the Portuguese data. The strongest relationships on the graphic are between GP3 and FP and between GP3 and FT. These relationships also happened in feedback messages written in Portuguese but with a weaker connection. We attribute this result to the fact that the English language feedback is much larger than that of the Portuguese. In other words, the English feedback provides high-quality information and makes the student reflect on his response. It is also possible to notice a good relationship between GP3 and GP5 and between GP3 and FS, which means that in addition to providing high-quality information, English language feedback also provides the student with praise to keep them motivated.

The most significant difference between the two graphs is the relationship with the GP6. In English, fewer feedback messages had this indicator. Thus, the instructor did not provide specific information for students to close the gap between current and desired performance. In contrast, most feedback messages had good practice GP3; consequently, high-quality information helped students understand where they went wrong and what the
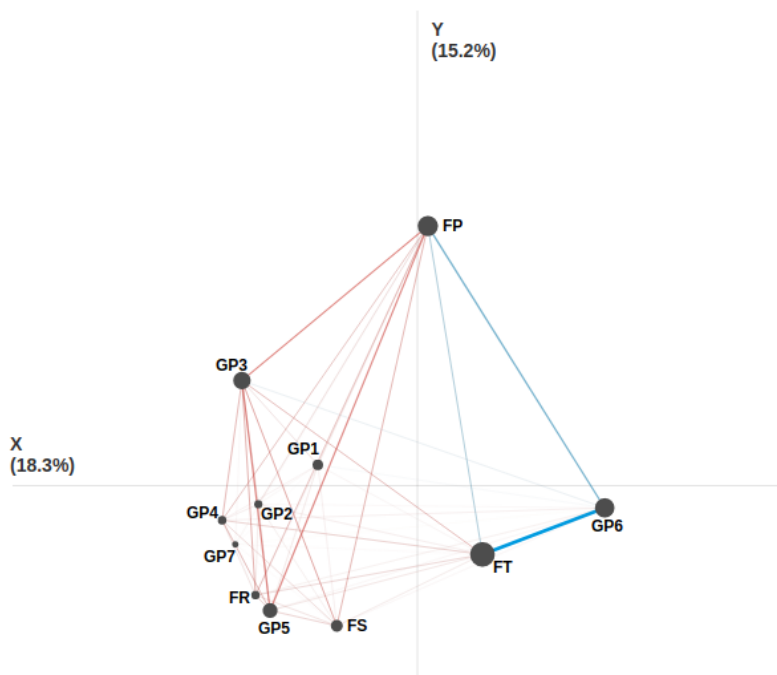
next steps in learning should be.

Figura 3. Subtraction between the Portuguese and English ENA network.

Figure 3 shows the subtraction of the two networks (EN and PT). The blue edges refer to the Portuguese language feedback and the red edges refer to the English language feedback. We can highlight in this network the clear division between Portuguese and English. The nodes and edges on the right are the most present in the Portuguese language feedback, that is, the FT, FP and GP6 good practice levels, highlighting the link between the FT level and the GP6 good practice, showing that there is a solid correlation between these two descriptors. Looking at the left side of the image, it is possible to see the predominant presence of the red color, that is, the feedbacks from the English language are more present in the FR, FS levels and in all the good practices, except the GP6 good practice.

6. Discussion

Answering our research question (Section 3), Figure 3 provided insights into the relationship between the feedback models evaluated. This graph shows that the FT and FP levels were present in many feedback for both languages. This result indicates the main goal of feedback messages was to show the student where they were wrong (FT level) and bring information that makes the student reflect on their response (FP level). In terms of good practices, the predominance of GP6 and GP3 demonstrates the effort to assist the students to close the gap between the desired performance and the activity. In both cases, the main indicators revealed by our analysis show and emphasize providing feedback on the activity (task and process), which is the most common type of feedback [Dunworth and Sanchez 2016, Al-Hamad and Mohieldin 2013, Maitra et al. 2018, Cavalcanti et al. 2019].

The results presented for the Portuguese language showed a strong connection between the FT and FP levels with good practice GP6. This connection shows that the instructors' main objective concerning feedback is to present where the student was wrong, suggesting to reflect and close the learning gap. However, the literature shows that it is also interesting to pass on other information that encourages self-regulation [Henderson et al. 2019]. In order to provide more information and motivation for students, the instructors should not focus only on task-related feedback [Wisniewski et al. 2020].

On the other hand, the English written feedback messages had a strong relationship between the FT and FP levels with the good practices GP3 and GP5. In other words, the English language feedback, in addition to showing where the student was wrong on their tasks, provide high-quality information with positive beliefs and self-esteem to keep students motivated. The average number can see one indicator of this multi-facet of English feedback of words (mean=437). An average is almost 15 times higher than that of feedback in the Portuguese language (mean=30).

It is essential to highlight that the feedback messages did not include many messages of FR and FS levels in both datasets. However, these levels of feedback could have a higher impact on students' self-regulation and satisfaction [Wisniewski et al. 2020]. The ENA charts clearly showed that there is a difference in the style of providing feedback between the two languages. That is, the levels and best practices most present in one language are not the most present in the other language. However, this difference could have occurred due to the feedback texts being extracted from different environments or courses or the difference in the number of messages between the two sets.

Finally, in summary, the main similarities between the two models were between the FT level and the GP6 good practice, the FP level and the GP3 good practice, and the FS and FR levels with the GP5 good practice, which shows, based on the analysis of this work, that there are relationships between the two models. A combination of the use of models by the teacher can be interesting, for example, to separate the feedback into different structures, with each structure being a level of feedback and within each level using the principles of good feedback practices that best suit.

7. Limitations

The main limitation of this work is the datasets. The Portuguese dataset had 1000 feedback messages obtained from only two different undergraduate courses (Biology and Literature). On the other hand, the English dataset has 272 feedback messages, a small amount to perform a fully comparative analysis. Thus, further analysis would be required to provide more insights on the topic. However, this paper focused on the potential of using ENA to compare different feedback models. In future work, we intend to evaluate a dataset with more feedback messages and different theories of good quality feedback using the same approach. The final goal is to develop a feedback quality measure using ENA that encompasses traditional models proposed in the literature.

Acknowledgement

Referências

[Al-Hamad and Mohieldin 2013] Al-Hamad, B. and Mohieldin, T. (2013). E-assessment as a tool to augment face-to-face teaching and learning environment. In *2013 Fourth International Conference on e-Learning Best Practices in Management, Design and Development of e-Courses: Standards of Excellence and Creativity*, pages 348–359. IEEE.

[Boud and Falchikov 2007] Boud, D. and Falchikov, N. (2007). *Rethinking assessment in higher education: Learning for the longer term.* Routledge.

[Cavalcanti et al. 2019] Cavalcanti, A. P., de Mello, R. F. L., Rolim, V., André, M., Freitas, F., and Gaševic, D. (2019). An analysis of the use of good feedback practices in online learning courses. In *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, volume 2161, pages 153–157. IEEE.

[Cavalcanti et al. 2021] Cavalcanti, A. P., Mello, R. F., Miranda, P., Nascimento, A., and Freitas, F. (2021). Utilização de recursos linguísticos para classificação automática de mensagens de feedback. In *Anais do XXXII Simpósio Brasileiro de Informática na Educação*, pages 861–872. SBC.

[Dunworth and Sanchez 2016] Dunworth, K. and Sanchez, H. S. (2016). Perceptions of quality in staff-student written feedback in higher education: a case study. *Teaching in Higher Education*, 21(5):576–589.

[Evans 2013] Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of educational research*, 83(1):70–120.

[Ferreira et al. 2018] Ferreira, R., Kovanović, V., Gašević, D., and Rolim, V. (2018). Towards combined network and text analytics of student discourse in online discussions. In *International Conference on Artificial Intelligence in Education*, pages 111–126. Springer.

[Gašević et al. 2019] Gašević, D., Joksimović, S., Eagan, B. R., and Shaffer, D. W. (2019). Sens: Network analytics to combine social and cognitive perspectives of collaborative learning. *Computers in Human Behavior*, 92:562–577.

[Gibbs and Simpson 2005] Gibbs, G. and Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and teaching in higher education*, (1):3–31.

[Hattie and Gan 2011] Hattie, J. and Gan, M. (2011). Instruction based on feedback. In *Handbook of research on learning and instruction*, pages 263–285. Routledge.

[Hattie and Timperley 2007] Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1):81–112.

[Henderson et al. 2019] Henderson, M., Ajjawi, R., Boud, D., and Molloy, E., editors (2019). *The Impact of Feedback in Higher Education: Improving Assessment Outcomes for Learners.* Springer International Publishing, Cham, Switzerland. Google-Books-ID: WyxQxgEACAAJ.

[Landis and Koch 1977] Landis, J. R. and Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, pages 363–374.

[Langer 2011] Langer, P. (2011). The use of feedback in education: a complex instructional strategy. *Psychological reports*, 109(3):775–784.

[Maitra et al. 2018] Maitra, S., Madan, S., Kandwal, R., and Mahajan, P. (2018). Mining authentic student feedback for faculty using naïve bayes classifier. *Procedia computer science*, 132:1171–1183.

[Matcha et al. 2019] Matcha, W., Gašević, D., Uzir, N. A., Jovanović, J., and Pardo, A. (2019). Analytics of learning strategies: Associations with academic performance and feedback. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pages 461–470. ACM.

[Mello and Gašević 2019] Mello, R. F. and Gašević, D. (2019). What is the effect of a dominant code in an epistemic network analysis? In *International Conference on Quantitative Ethnography*, pages 66–76. Springer.

[Nicol and Macfarlane-Dick 2006] Nicol, D. J. and Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2):199–218.

[Osakwe et al. 2021] Osakwe, I., Chen, G., Whitelock-Wainwright, A., Gaševic, D., Cavalcanti, A. P., and Mello, R. F. (2021). Towards automated content analysis of feedback: A multi-language study. In *Proceedings of the 14th International Conference on Educational Data Mining*.

[Parikh et al. 2001] Parikh, A., McReelis, K., and Hodges, B. (2001). Student feedback in problem based learning: a survey of 103 final year students across five ontario medical schools. *Medical education*, 35(7):632–636.

[Rolim et al. 2019] Rolim, V., Ferreira, R., Lins, R. D., and Găsević, D. (2019). A network-based analytic approach to uncovering the relationship between social and cognitive presences in communities of inquiry. *The Internet and Higher Education*, 42:53–65.

[Sadler 1989] Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, 18(2):119–144.

[Sadler 2010] Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5):535–550.

[Shaffer 2017] Shaffer, D. W. (2017). *Quantitative ethnography*. Cathcart Press, Madison, WI.

[Shaffer et al. 2016] Shaffer, D. W., Collier, W., and Ruis, A. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3):9–45.

[Shaffer et al. 2009] Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A. A., and Mislevy, R. (2009). Epistemic Network Analysis: A Prototype for 21st-Century Assessment of Learning. *International Journal of Learning and Media*, 1(2):33–53.

[Wisniewski et al. 2020] Wisniewski, B., Zierer, K., and Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10:3087.