Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal
PEER REVIEWED PAPERS
LEARNING ANALYTICS: FOR A DIALOGUE BETWEEN TEACHING PRACTICES AND EDUCATIONAL RESEARCH
Vol. 15, n.3, 2019
ISSN: 1826-6223 | eISSN: 1971-8829

USER RATING AS A PREDICTOR OF LINGUISTIC FEEDBACK QUALITY IN QUESTION AND ANSWER PORTALS

Simone Torsani

University of Genova, Italy
Simone.torsani@unige.it

Keywords: Question and Answer, informal language learning, corrective feedback, Italian

Question and Answer portals allow users to post and answer questions on different issues, among which foreign languages. The present paper focuses on feedback requests, i.e. questions in which users of the site ask for linguistic feedback on short sentences or phrases. In particular, it reports on a research on the degree of reliability of the evaluation of answers provided by the portal's users to identify correct and good linguistic feedback. An observational approach was adopted for about 600 answers in the Italian version of Yahoo! Answers. Each feedback was evaluated by two expert teachers and their rating was then compared with the evaluation provided by the site's users. Results show that, while the correlation between the votes of the community and the rating of the experts is rather weak, answers with a positive evaluation generally contain a correct feedback. We conclude, therefore, that caution must be exercised when utilising users' evaluation as guidance on feedback choice.

for citations:
Torsani S. (2019), *User rating as a predictor of linguistic feedback quality in Question and Answer portals,* Journal of e-Learning and Knowledge Society, v.15, n.3, 227-237. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1135045

1 Introduction

The present observational research aims at investigating whether user assessment of answers in a Question and Answer portal (hence, Q&A) is a reliable marker of the quality of linguistic feedback provided therein.

Everyday many learners autonomously resort to online services, which constitute one among the many options nowadays available for language learning. Although such venues generally operate outside the domain of formal education, it is nonetheless important for language educators and advisors to know their potential to help learners make the most of them. In particular, since the format of Q&A implies that a question receive different answers, it is important for the learners to develop strategies to identify the most suitable one(s).

1.1 Q&A and informal language learning

Since the Internet has been available to the public, experts have been prompt to recognize its potential for language learning. Interaction has been pivotal in language learning theories for about the last forty years and it comes as no surprise that the possibility to interact online, especially with native speakers, has ever since been seen as a great benefit for learners (see e.g. Chapelle, 2006; Ziegler, 2016). This trend has gained momentum with the rise of the so-called web 2.0 and social media, in which the role of users further expanded to that of users and producers of content (Harrison & Thomas, 2009). Research on social media and language learning has consequently flourished in the last years with many works investigating the potential of such tools (see e.g. Lomicka & Lord, 2016; Zheng, Yim & Warschauer, 2018). The present research focuses on Social Networking Services (hence, SNS), in particular the use of SNS for self-directed, informal learning, i.e. the purposeful usage of SNS to learn or to improve a language (Reinhardt, 2019).

Although quite neglected by second language research, Q&A (e.g. Yahoo! Answers or Quora) have much to offer to learners. The mechanism behind Q&A is rather simple: a user posts a question and other users provide an answer and/or, in some cases, evaluate answers by other users (see Adamic *et al.*, 2008 and Jin *et al.*, 2015 for an overview of Q&A). Both activities, namely providing and assessing content, constitute the backbone of social media. Since a feature of Q&A is that questions receive different answers, users' evaluations of answers have a central role in the economy of these services. Indeed, as is common in online venues (for instance, in online commercial sites), users may rely on such evaluations for help on what to choose, in this case the most suited answer to their question.

1.2 Corrective feedback in Q&A

Starting from general-purpose taxonomies of questions in Q&A Torsani and Dettori (2018) argue that this format yields to different language-related usages. However, while they recognise that each of such usages may influence language learning, it is to what they call "language support" questions that they look to as a remarkable option for language learning. Language support questions focus on such issues as grammar rules, vocabulary or feedback requests and their answers generally provide linguistic material learners can process and hence improve their linguistic skills. Asking such questions, in either a formal or informal environment or fashion, is a common experience for language learners and Q&A simply amplifies the number of potential experts. Among language support questions, feedback requests constitute a promising subset because learners can ask experts or native speakers for a fast and informal linguistic feedback on their utterances. Not a secondary asset for them.

While feedback has been traditionally researched from the perspective of classroom teaching and learning (see e.g. Brown, 2016 and Lyster & Saito, 2010), the spread of network technologies has meant a broadening of interests towards peer feedback delivered in online interaction (see e.g. Bower & Kawaguchi, 2011 and Vinagre & Muñoz, 2011). However, in the case of SNS learners have raised concerns about the quality of the feedback provided by peers (Stevenson & Liu, 2013). Dispelling any such concern, therefore, is of primary importance for learners and advisors alike, in order to assess whether these services deserve a position among the tools for language learning.

1.3 Investigating feedback in Q&A through Learning Analytics

In Q&A a request request receives multiple answers and the questioner must choose the one that best fits their needs. This leads to an important issue: how can a learner be helped choose the best answer? As stated before, the evaluation of an answer provided by other users should ideally constitute a reliable tool for learners. A premise of social media is indeed what is known as the "wisdom of the crowds" (Surowiecki, 2005), best exemplified by Galton's experiment in which the mean of all the estimates of the weight of an ox was close to the real weight of the animal (Galton, 1907). This fascinating perspective, in high regard in the heyday of social media and Web 2.0, has however been progressively questioned, as social media have in some cases become a channel for unscientific information (see e.g. Vosoughi, Roy & Aral, 2018). The issue of users' evaluation validity, therefore, arises also from an educational perspective and its role, in this case, in helping a questioner choose the best feedback must consequently be submitted to scrutiny. Learning Analytics (hence, LA) appear to be a convenient tool to achieve an understanding of this issue. In particular, given the social nature of Q&A, it is to Social Learning Analytics (hence, SLA, Shum & Ferguson, 2012) we turn to for this task (see below).

1.4 Research question

In line with the premises of social media, we expect users' evaluations to be a reliable indicator of the quality of an answer. We also expect users' evaluations to be indicators of good and bad feedback alike (i.e. negative user assessment indicates a bad feedback and positive user assessment indicates a good feedback). Finally, based on the notion of wisdom of the crowds, we expect such reliability to increase with the number of votes assigned to an answer. Therefore, the present research aims to answer the following question:

1. Are users' evaluations of answers a valid means to help a questioner choose good linguistic feedback? In particular:
    - Is there a correlation between the overall users' evaluation of an answer and its quality?
    - Are user ratings more reliable when detecting good or bad feedback?
    - Does reliability increase with the number of ratings?

2 Materials and methods

2.1. Methodological Approach

Because of the social nature of Q&A we have adopted a SLA approach (see above), proposed by Shum & Ferguson (2012), who set off from the features (both technical and ecological) of social media for learning to define a peculiar ambit for LA. Such approach focuses on the participatory nature of online social learning rather than on the features of formal education. As they argue, «the focus of social learning analytics is on processes in which learners are not solitary, […] but are engaged in social activity, either interacting directly with others (for example, messaging, friending or following), or using platforms in which their activity traces will be experienced by others (for example, publishing, searching, tagging or rating)» (p.5). Users' ratings constitute one of the types of data SLA takes into account. In particular, the analysis of any kind of content produced by participants falls within the "content analytic" category of their proposed taxonomy (Ferguson & Shum, 2012), which consists in the application of LA principles to user-generated content. Because the purpose of LA and SLA alike is to provide information to improve teaching and learning, such approach is important in that it can guide potential learners towards using answer evaluations as a reference point. Because of the explorative nature of the present research, we adopt here a somewhat simple statistical approach to investigate whether users' evaluations of answers containing linguistic feedback are a reliable predictor of such quality.

2.2. Data set

A hundred feedback requests and the corresponding 614 answers were collected from the Italian version of the Yahoo! Answers portal. To be included in the data set, questions needed to:

- be posted by a learner of Italian;
- request feedback on a phrase or sentence;
- contain at least one overt error;
- have at least one answer;

Questions and answers were collected in a spreadsheet, in which every row contained a single answer together with the corresponding question and users' evaluation; for instance (all texts from the data set are reported as they are with no correction):

> [question id] 20130127122439AALCNCL; [question title+body]: *quale frase e' giusta?? (in italiano)? si dice1. mi piacciono tutti i lavoro che riguardano l'italiano 2.mi piacciono tutti i lavoro che riguardano all'italiano???? grazie in anticipo sono straniera...;* [answer]: *Mi piacciono tutti i lavori che riguardano l'italiano. Così e giusta*: [positive evalutations] *1*; [negative evaluations] *0*;

2.3 Research design

Two mother tongue teachers of Italian (hence, the Experts) independently rated each answer with a holistic score ranging from -5 to 5 on a version of the above-mentioned spreadsheet from which users' evaluations were removed. Questions in the data set were not chosen based on factors such as difficulty or frequency of errors and are, consequently, quite heterogeneous in this respect. Therefore, the Experts received no assessment grid or specific instruction as to how rate answers: they were only asked to rate the quality of the linguistic feedback based on the request.

We then calculated the mean of the two scores as a reference score (hence, Expert Assessment, EA) for each answer against which we compared the assessment of the portal's users, i.e. the difference between the sum of positive and negative assessments (hence, UA). In the example quoted above the answer has positive evaluations =1; negative evaluations =0; and, consequently, UA =1. We excluded zero values from the number of UA either because the answer received no assessment or because they had an equal number of positive and negative evaluations. UA and EA are different measures. UA corresponds to the sum of all individual (negative and positive) votes. A user can only say whether she/he approves or not an answer, without specifying how much. EA, on the contrary, corresponds to the mean of two scores given on a scale; in other words, experts can specify if they find an answer particularly good or bad.

To answer the research question(s), different tests were run to measure the agreement between experts and users.

First, the correlation between EA and UA was calculated to determine whether UA can be considered a good marker for answer quality, i.e. the larger the overall UA the higher EA. We expect that the better the answer (receiving a high score from the Experts) the higher number of positive votes it receives by the users.

To observe whether users are more capable of detecting good or bad feedback, we ran a chi-square test considering the number of positive/negative UA and their agreement with positive/negative EA. In this case, we adopted a binary perspective; votes were considered in agreement if they shared the same overall positive or negative orientation.

Finally, to observe whether the ability to detect good/bad feedback increases with the number of evaluations, we ran a chi-squared test on the number of agreements and non-agreements on answers receiving one, two/three and four or more evaluation. Here, we assume that the higher number of votes an answer receives, the higher the agreement with the experts.

3 Results and discussion

The Experts provided 500 (81.43%) overall positive and 114 (18.57%) negative ratings (*N*=614). As both explained, they independently adopted a rule of thumb according to which a simple, but correct, feedback would receive a small positive score (1 or 2); a good feedback (e.g. one comprising a useful linguistic focus on the error) would receive a higher value; a useless one (e.g. an off-topic answer) would receive 0 or -1; finally, a misleading one (i.e. a feedback which does not correct errors) would receive a score below 0. The figures reveal that many instances of feedback were acceptable to them and a large share also good to excellent. According to the Experts, therefore, about 4/5 of the answers provide a (more or less) useful feedback, which is what one may reasonably expect since answers consist in feedback on the respondents' mother tongue.

The members of the community provided 928 individual votes (1.51 votes per answer): 478 positive and 450 negative ones. In our data set 214 (35%) answers have an overall positive and 166 (27%) a negative UA score. 193

(31%) answers received no evaluation, while 41 (7%) answers received an equal number of positive and negative evaluations (hence referred to as neutral), thus resulting in UA=0.

(RQ 1.a) A correlation test between UA and EA returned a significant, but quite weak, positive result. For this test, only answers with at least one user vote were considered. A Pearson correlation of r(*N*=422)=0.31, p<0.01 was found between UA and EA. UA scores explain R2=9% of the variance of EA. UA is an indicator, albeit weak, of the quality of an answer.

Next, we focused on the agreement of the overall positive or negative sign between EA and UA.

Table 1
UA AS A PREDICTOR OF EA

	Correctly identified by UA	Not or incorrectly identified by UA	total
Positive EA	197 (39.40%)	303 (60.60%)	500
Negative EA	48 (42.10%)	66 (57.90)	114

Table 1 reports all the cases in which UA correctly identifies or not feedback, i.e. UA and EA have the same (positive or negative) sign. Only 39% of correct/good instances of feedback received a positive UA, while 61% received a negative, neutral or no evaluation. A similar ratio (42% vs. 58%) applies to negative EA.

Table 2
AGREEMENT BETWEEN UA AND EA

	Agree with EA	Not agree with EA	total
Positive UA	197	17	214
Negative UA	48	118	166
total	245	135	380

(RQ 1.b) While most answers are not correctly identified through UA (which confirms the scarce correlation between UA and EA), a clearer picture emerges if positive and negative UA are considered separately. Table 2 reports the number of instances of overall positive and negative UA and the cases in which these agree or not with EA. A chi-square test on agreement returned a strong result, with $\chi 2$ (1, *N*=380) = 162.71, p<0.001: positive UA were in line with EA, while negative evaluations generally were not. Therefore, while not all positive EA are identified through UA, a positive UA generally entails a positive EA. This scenario, however, does not apply to negative UA. In other words, an overall positive UA is a good predictor of the correctness of the feedback contained in the answer, while an overall negative one is not a good predictor of a bad answer.

(RQ 1.c) A chi square test was run to determine whether the ratio between agreement/non agreement changes as the number of evaluations increases, but the result was not significant and it is therefore not possible to reject the null hypothesis that agreement does not change based on the number of assessments (see Table 3).

Table 3
AGREEMENT BETWEEN EA AND UA BASED ON THE NUMBER OF VOTES

Number of votes	Agree with EA	Not agree with EA
1	121	64
2/3	88	52
4 or more	36	20
Total	245	136

The answer to our research question was not as straightforward as expected and our findings suggest that, while user assessment can provide some support for learners in choosing a good feedback, caution must nonetheless be exercised.

First, a significant correlation does exist between UA and EA, but it is rather weak and is of little help in assessing the overall quality of feedback. This was a major expectation, since we assumed that the better an answer, the higher the number of positive evaluations. However, this is not always the case and factors other than linguistic correctness of a feedback must intervene in the evaluation of an answer on the part of the members of the community. Politeness, for instance, seems to be rather important for some users, who sometimes assign negative ratings to an answer when it contains offensive or apparently impolite language regardless of the correctness of the feedback. For instance, q.id 20100827102957AA28mFP asks which of two forms is correct: (…) *amore mio senza di te muorirei... il mio dilemma è : muorirei oppure morirei?* (my dear, I would die without you… what I do not know is morirei or muorirei?), a user correctly answers morirei, but somehow awkwardly adds “Italian (here meaning grammar) is not an optional”. While the Experts based their assessment on the correctness of the feedback and gave this answer positive evaluations, users gave it six negative votes (and no positive one), thus resulting in a strongly negative UA.

The answers to our second and third sub question are perhaps more encouraging. Although most instances of feedback (about 60%) are not correctly identified by UA, a positive UA generally entails a correct/good feedback. Ideally, a learner should aim at the best answer; however, since their objective is receiving linguistic feedback, even a simply correct answer constitutes a useful support. Furthermore, a connection between number of votes and ability to detect good/bad feedback could not be demonstrated and it is not possible to reject the null hypothesis that the two are unrelated. If this were confirmed it would mean that, counter to the mythology of social media, even a single positive vote is a good marker of correct feedback.

A somewhat positive balance can finally be drawn from these findings. Indeed, since feedback quality in SNS is a concern for learners (Stevenson and Liu, 2013), our findings demonstrate that, in the case of our data-set, user votes constitute a valid support in identifying good feedback.

Conclusions

The present research has focused on feedback delivered through Q&A portals and, in particular, on the reliability of users' evaluations of answers. The findings show that, while there is a certain discrepancy between experts and users, user ratings constitute a reliable tool for detecting correct/good feedback.

From the vantage point of language education, the integration of feedback through Q&A has different implications, of which we will focus here on the impact of the findings of the present research from a LA perspective. Since the main objective of LA is to provide information to improve learning, in this case informal, our findings suggest that, with due caution, feedback in Q&A can be a valid option for learners, who can rely on users' vote when they need to choose an answer. In a survey of SNS for language learning, Lin, Warschauer and Blake (2016) found that receiving feedback from peers was much valued by their participants. However, while participation in the services considered in that study suffered from a somewhat sharp drop in the long run, Q&A constitute a fast and lightweight alternative, which can be more easily integrated into everyday language-related formal and informal activities.

While it furthers our knowledge of informal language learning in SNS, the present research has, however, some limitations, which should be kept in mind when considering its findings. The first limitation of this research is that it focuses on the correct/incorrect dichotomy and does not account for the factors affecting users votes: for instance, in the discussion we hinted at the possible influence of affective factors in determining users' votes. The scant number of user votes constitutes the second limitation. When considering user evaluations in other Internet services (e.g. feedback on products in e-commerce sites) figures are considerably higher, therefore our findings should be confirmed by research on more sizeable data-sets. A third limitation is that the research was conducted on the Italian language and it is not clear whether its findings are generalizable to other languages. While, for instance, in our data set it was arguably native speakers that provided feedback and votes, in the case of more diffused languages, like English, also non-native (and non-proficient) speakers might participate and alter the overall quality/assessment balance.

Besides these limitations, however, we must acknowledge that even a narrow ambit like feedback in Q&A appears to be a rather complex phenomenon and different aspects must be taken into account when trying to provide an accurate picture of it. For instance, we did not focus on fundamental issues, such as the ability of the learners to recognize (and choose) the best answer and the contribution of users' votes to this choice. As the case of affective factors seems to suggest, feedback evaluation in Q&A stretches beyond the correctness of the answer. Both the quantitative and qualitative perspectives of SLA illustrated in Ferguson & Shum (2012), therefore, offer important insights in this matter and their findings should be integrated to achieve a clearer understanding of this tool.

REFERENCES

Adamic, L. A., Zhang, J., Bakshy, E., & Ackerman, M. S. (2008, April). Knowledge sharing and yahoo answers: everyone knows something. In *Proceedings of the 17th international conference on World Wide Web* (pp. 665-674). ACM.

Bower, J., & Kawaguchi, S. (2011). Negotiation of meaning and corrective feedback in Japanese/English eTandem. *Language Learning & Technology*, *15*(1), 41-71.

Brown, D. (2016). The type and linguistic foci of oral corrective feedback in the L2 classroom: A meta-analysis. *Language Teaching Research*, *20*(4), 436-458.

Chapelle, C. A. (2006). Interactionist SLA theory in CALL research. In *CALL research perspectives* (pp. 65-76). Routledge.

Galton, F. (1907). Vox populi (the wisdom of crowds). *Nature*, *75*(7), 450-451.

Ferguson, R., & Shum, S. B. (2012). Social learning analytics: five approaches. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 23-33). ACM.

Harrison, R., & Thomas, M. (2009). Identity in online communities: Social networking sites and language learning. *International Journal of Emerging Technologies and Society*, *7*(2), 109-124.

Jin, J., Li, Y., Zhong, X., & Zhai, L. (2015). Why users contribute knowledge to online communities: An empirical study of an online social Q&A community. *Information & management*, 52(7), 840-849.

Lin, C. H., Warschauer, M., & Blake, R. (2016). Language learning through social networks: Perceptions and reality. *Language Learning & Technology*, *20*(1), 124-147.

Lomicka, L., & Lord, G. (2016). Social networking and language learning. *The Routledge handbook of language learning and technology*, 255-268.

Lyster, R., & Saito, K. (2010). Interactional feedback as instructional input: A synthesis of classroom SLA research. *Language, Interaction and Acquisition*, *1*(2), 276-297.

Reinhardt, J. (2019). Social media in second and foreign language teaching and learning: Blogs, wikis, and social networking. *Language Teaching*, *52*(1), 1-39.

Shum, S. B., & Ferguson, R. (2012). Social learning analytics. *Journal of educational technology & society*, *15*(3), 3-26.

Stevenson, M. P., & Liu, M. (2013). Learning a language with Web 2.0: Exploring the use of social networking features of foreign language learning websites. *CALICO journal*, *27*(2), 233-259.

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Torsani S., & Dettori G. (2018). Una tassonomia di domande di argomento linguistico nei portali di Domande e Risposte. *Ricognizioni 5*(10), 81-96.

Vinagre, M., & Muñoz, B. (2011). Computer-mediated corrective feedback and language accuracy in telecollaborative exchanges. *Language Learning & Technology*, *15*(1), 72-103.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146-1151.

Zheng, B., Yim, S., & Warschauer, M. (2018). Social media in the writing classroom and beyond. *The TESOL Encyclopedia of English Language Teaching*, 1-5.

Ziegler, N. (2016). Synchronous computer-mediated communication and interaction: A meta-analysis. *Studies in Second Language Acquisition*, *38*(3), 553-586.