A Novel Tool for Online Community Moderator Evaluation

Alicia J. W. Takaoka[1,2(✉)] [ID]

[1] University of Hawai'i at Hilo, Hilo, USA
ajwilson@hawaii.edu
[2] University of Hawai'i at Mānoa, Honolulu, USA
https://cms.uhh.hawaii.edu/faculty/atakaoka/

Abstract. This study introduces a new instrument for leadership evaluation in online forums and other online communities which was developed using a grounded approach. Questions that emerged from the literature were then evaluated to create hypotheses that guided the development of an instrument for moderator evaluation. The Moderator Evaluation Contingency Scale (MECS) is modified from Fiedler's contingency model to determine if a moderator is more task- or relationship-oriented in his or her approach to moderation and interactions with other members of a community. The MECS was developed and tested on Reddit in 2013–2014 using random sampling for Forum selection, moderator selection, and interactions with users. A content analysis using the MECS to evaluate posts was found to be a viable measure of a moderator's ability to perform tasks like removing content as well as his or her ability to interact with users. Bots were analyzed using the MECS as well to determine bias. Next steps include making the instrument available for use by social media and niche community sites, administrators, and other moderators.

Keywords: Moderation · Online community · Online leadership · Fiedler's contingency model · Interaction instrument · Leadership evaluation

1 Introduction

Leadership is well-studied in academic literature. Studies across disciplines overlap and converge by observing the behavioral phenomena in different settings to evaluate qualities, like perception and emotion, of the ideal leader in a given situation [15]. This study updates Fiedler's contingency model of leadership to evaluate moderators and content moderation in online forums and communities. While Fiedler's model is not without its limitations [26], the model is versitile and adaptable for the both evaluating the needs of a community as well as evaluating existing work of a potential moderator. Researchers have been examining how to effectively lead teams in virtual settings (e.g. [5,11,25,28,30,31], and [35]), and leading in virtual classrooms [33]. Still most of the research is about leadership in the context of business and face-to-face interactions rather than

online interactions. Of the online leadership that is being studied, it relates to business. The body of work on online leadership is growing, but the research conducted on the behaviors of online leaders in forums and other types of social media is still minimal.

2 Background

This study evaluates the ability of measuring leadership styles based on moderator interactions, reminiscent of Vlachopoulos and Cowan's study of online moderation [34] before they become moderators. While this study does not assume a community informatics approach, it acknowledges the need for an administrator to evaluate a potential online moderation leader based on the needs of the group, forum, or community. This assumes that administrators will know the type of leadership style needed based on content and community composition [19].

Leadership is important in online spaces. Avolio and Kahai with others [2,3], and [4] have advocated the move to e-leadership for businesses in a current context [12] as well as re-evaluated the state of e-leadership studies ten years later [6]. Cosley et al. and others have stated that both oversight and moderation are necessary in online communities. Even if the behavior persists in the form of griefing, trolling, or becoming an archetype (e.g. [10,14,20], and [27]), moderation is a function of the site that is trusted to "key members who manage and maintain the community. These tasks include moderation, governance, welcoming new members, and building Frequently Asked Question (FAQ) lists" [10]. These functions are what separate some leaders from everyday users. There are several leadership styles that can be identified, but this paper focuses solely on Fiedler's contingency model of leadership.

2.1 Leadership Styles

According to Lewin, there are three primary leadership styles. They are authoritarian, participative [24], and laissez-faire [8]. All three styles operate effectively on Reddit in various subReddits dependent on different situational contexts. Authoritarian leadership is exhibited when the leader has very clear expectations and outlines how and when actions should be taken. Authoritative leadership is best used in situations where there is little time for group decision-making, as is the case when a moderator must choose to remove a post or, in some cases, ban a user. In this style of leadership, there is a clear division between the leader and the participants [29]. On Reddit, moderators have dual roles as both a user and as a moderator. When they participate in a conversation, however, the role is clearly visible. The moderator's name is green, and there is at least one insignia next to their name: [**M**].

2.2 Fiedler's Contingency Model

Fiedler's contingency model of leadership was developed in the 1960s to determine if a person is more task-oriented (task motivated) or concerned with inter-

personal relationships (relationship motivated) when making leadership decisions. Fiedler states, "a leader's performance is contingent on two interacting factors: (a) the individual's goals and needs or motivational structure and (b) the degree to which the situation provides the leader with control and influence over the outcome of the task" [18]. The instrument used to measure a leader's motivation is the least preferred coworker scale. A current leader is given this scale, which uses polar opposite adjectives, to describe his or her least preferred coworker. Leaders with lower scores who use mainly negative words to describe the least preferred coworker tend to be more task-oriented with leaders who have higher scores and use positive words to describe the least preferred coworker are more relationship-oriented. From this, the following question was developed: *Can Fiedler's leadership instrument be modified to evaluate online community moderators?*

2.3   Moderation

Moderation is important in online communities for many reasons. Administrators, designers, and community members have to regulate the content in the site in some way to ensure the quality of posts, the relevance of the topic, and that users interact in a way that is respectful or beneficial to one another. Grimes et al. explain that terms of use are usually obscured to the user by legal language, but they should be created by the users [22]. In this way, users can create a sense of ownership in the site and are aware of the policies. One way of making policies clearer to users are to have visible leaders that can be relied upon to explain why rules are in place and maintain a respectful environment.

Moderation is both a universal concept that transcends one group, yet a moderator is situated in a specific structure and culture. They are seen as leaders and figures of authority. Users look to moderators when other users break the rules of the community for punishment and resolution of a situation. In some online communities, users even look to moderators to facilitate discussions. This study examines whether content moderation is function of leadership styles. Specifically, this study looks at content moderation as a function of leadership determine if a moderator can switch between leadership styles depending on the situation.

2.4   Building Trust and Social Roles

The success of any online community is dependent on trust. Trust must be built between administrators and others in leadership as well as among the users who belong to the community. In some ways, this is done by users building an identity, creating a profile, and interacting with one another based on mutual friends or similar interests [9, 16, 23]. One way to build trust is through effective moderation of a site. Since a subReddit is a niche community with rules unique to that community, effective moderation is important for both new and existing users of those subReddits.

Moderation is a social role that a user can assume. It is a promotion in some ways, and a burden in others. Social role theory says that our behaviors, actions, thoughts, and wants are prescribed by a specific set of socially determined roles. Roles are the parts that we play, and these roles vary depending on the situation. One does not play the same role when she is a coworker as when they are wife, mother, or sister or husband, father, or brother. Each role is unique to a specific set of situations. The social construction of roles is maintained as we transition to interactions in online communities. Gleave, Welser, Lento, and Smith look at the foundation of a social role. The authors believe a role is rooted in and can only be understood from a context of structure and culture. The authors write:

> Our definition asserts that social roles begin from a structural foundation in simple commonalities in behavior...the role of father fills certain basic social needs and is therefore recognizable across cultures, both by outside observers and individuals living within a given social environment. Similarly, many social roles, especially those that are newly emerging, will have distinctive social structural foundations even if they have not yet developed the same level of recognition within and across cultural settings [21].

Moderation is both a concept that transcends borders and is unique to the specific community in which moderation occur. From this, the following question was developed: *Can an instrument be used to evaluate current and potential leaders according to the needs of the online community?*

2.5 Reddit

Reddit is an online community in which users post links to news and entertainment. A podcast derived from Reddit, called Endless Thread, can be heard on NPR. As Reddit is a collection of posts, it can be considered a bulletin board community, one where the main source of contribution and primary purpose is the posting of content, whether created or shared by links, to be discussed by users. As Anders writes, "Reddit is a giant bulletin board made up of 185,000 active forums with their own obsessions: science trivia, political arguments, video game critiques, jokes and photos—lots of photos, of which more than a sprinkling are of cats or naked women" [1]. Reddit has subReddits, created and managed by Redditors (any user belonging to the Reddit community), that boarder on illegal as they are extreme. National Socialism and Cocaine are two examples of extreme subReddits. Reddit gets over 3.4 billion page views a month. It was purchased by Advance Publications, which is also the parent company for Conde Nast, in 2006.

One of the appeals of Reddit is the anonymity afforded to users. Users can create as many usernames associated with a single account as they want, and they can change their names an infinite number of times. This is a freedom limited in most social media and niche communities.

3   Methods

This sequential study led to the development of an instrument that can be used by moderators to evaluate potential moderators in their online communities. The first phase was to determine the overall leadership style of Reddit and subreddits. Next was to determine the translatable dichotomous characteristics from the Least Preferred Coworker Scale for online moderation. Next was to identify moderators and their posts in order to test the instrument.

Ten Moderators were selected at random from the twenty main forums, and ten posts each moderator, or 100 total posts, specifically moderating another user were selected to evaluate the effectiveness of the dichotomous pairs. These posts include comments, acts of moderation, stating the rules of the subReddit, deleting posts or comments, and banning users. This data was scaled using Fiedler's Contingency Scale. Categorical analysis was used to evaluate the level of interaction between the moderator and Redditor receiving the moderation. The dichotomous pairs adapted for the MECS canbe seen in Fig. 1. This study was repeated with a larger data set. Twenty moderators were selected, and they moderated a total of 27 subreddits between them. Their posts, 270 in total, were selected across subreddits.

Fig. 1. Categories adapted from Fiedler's Contingency Model- Least Preferred Coworker Scale.

The scale for analysis was reduced from 8 points to 5 points for categorical simplicity. Moderated posts were organized by moderator and post number in the order the content was retrieved from Reddit, as can be seen in Fig. 2. Each post was evaluated manually based on the moderated interaction, denoted by [**M**] only. Further interaction between the moderator and Redditor was only included in the content analysis if the interaction was relevant to the moderation. In other words, interactions that were irrelevant to the moderation- like conversations with other Redditors, images, and phatic communication- were not included in analysis. Response time was excluded but is available in the instrument, as seen in Fig. 4.

4   Results

Moderators employ a wide range of leadership styles. Figure 3 highlights this by displaying the total scores and averages of each moderator evaluated. Six moderators were task-oriented, thirteen were flexible or neutral, and only one

Post	Mod#	Timely\ Untimely	Unfriendly\ Friendly	Hostile\ Supportive	Unkind\ Kind	Guarded\ Open	Meant to Discuss?	Total	Notes
1	10		3	3	3	3	1	13	explanation
2	10		3	3	3	3	1	13	explanation
12	10							0	delete
4	10		1	1	1	1	1	5	
5	10		1	1	1	1	1	5	
6	10		2	2	2	2	1	9	
7	10		2	2	2	2	5	13	
8	10							0	post- gives rights to all redditors
3	10		2	2	2	2	1	9	
10	10							0	NA
11	10		3	3	3	3	1	13	
9	10		3	3	3	3	1	13	explanation
	10		20	20	20	20	13	93	10.33333333

Fig. 2. Example of the MECS criteria for evaluation in action.

moderator was relationship-oriented. Of the twenty moderators analyzed for this study, two were bots. As expected, their scores were neutral across the board. Their programmed responses were not meant to incite discussion. Both bots averaged a score of 15, according to the MECS. Scores ranged from 5.6 on the low end to 23 out of 25 as the highest. These scores accentuate different preferences for moderation as well as shape the role of the moderator in and the culture of a given subreddit. The given ranges for MECS overall scores and a moderator's leadership ability can be defined as:

– 5.00–11.99: Task-Oriented
– 12.00–18.99: Flexible to Neutral Moderator
– 19.00–25.00: Relationship-Oriented

Relationship-oriented moderators excel at reaching the user. They are willing to give users chances to modify behavior and are compassionate of users for errors like posting in the wrong place or breaking a rule once or twice. These moderators can acculturate new users to a community efficiently because they are less strict and more patient than task-oriented moderators. The median range is the flexible or neutral moderator. A flexible moderator is able to judge a situation and act accordingly. When the situation requires a moderator to acculturate or explain a situation to a user, the flexible moderator is able to do this. This is in contrast to [19] found and thus needs more exploration using MECS.

Moderators who have a lower score tend to be more abrupt in their interactions with users. They deal well with tasks but may lack in communicating meaningful information to the user. Moderators with higher scores tend to have better interactions with users. These moderators can successfully explain a situation or action with tact. It is worth noting that no person scored a 15. Only bots achieved a score of total neutrality by the crafted message that is sent to a user in the comments section. This message refers users to contacting a person if they have additional questions about the moderation.

The bots had neutral scores of 15. In all cases of a bot's presence, the responses were courteous, informative, provided instruction on either the removal

Mod#	Response Time	Unfriendly\ Friendly	Hostile\ Supportive	Unkind\ Kind	Guarded\ Open	Meant to Discuss?	Total	MECS Score
13		33	32	33	33	24	155	5.61
6		23	23	23	25	13	107	8.33
24		19	22	20	21	25	107	9.42
10		21	20	21	27	13	102	10.33
12		13	13	13	14	15	68	10.36
25		33	33	32	34	26	158	10.57
17		21	21	21	24	20	107	12.15
5		8	8	8	8	4	36	13
18		31	29	29	29	24	142	13
26		19	21	21	21	15	97	13.29
7		46	45	45	47	16	199	14.41
22		48	48	48	58	32	234	14.41
4		37	37	35	53	14	176	15
20		51	51	51	51	51	255	15
9		26	25	27	37	15	130	15.91
11		24	24	24	25	26	123	16.18
2		39	39	39	65	13	195	17.23
23		27	29	26	27	33	142	17.55
27		52	46	49	55	15	217	18
8		11	10	12	20	4	57	23

Fig. 3. MECS analysis of all moderators.

of content or comments about the post, and included disclosure of being a bot. In the disclosure, instructions for following up with a live person were also included. The same bot had different language written into its response for another type of violation. The response is less cordial and contains stronger language. The response includes the disclaimer to contact the moderators if the user has questions, but the tone is different, showing that users have pre-programmed responses containing different tones by using bold typeface to highlight matters of importance so that the bot still demonstrates the ability to be flexible depending on the severity of the moderation. However, a bot is still not as versatile in its moderation capabilities as a person. Each bot can only address a limited selection of violation types, and a person must continually check and update the code for accuracy.

5   Discussion

This study has found that it is possible to adapt an existing leadership mechanism for online communities and forums, and moderators can be evaluated from the third-person perspective based on the needs of a specific community. Moderators have different functions based on the needs of specific forums. This study has shown that relationship-oriented moderators are harder to find than task-oriented moderators, but more forums require task-oriented moderators than relationship-oriented moderators. However, more widespread testing of the instrument across subreddits and niche communities is required.

Moderators who fell into the flexible to neutral category with a score of 14 or higher on the MECS were able to positively interact with users during moderation. Moderators with a score below 14 tended to be more hostile or abrasive in their moderation style and interactions with their community of users. Although past interactions among moderators and the same users were unknown, the appearance of foul language indicated hostility on the MECS scale. It is possible to objectively review potential moderators based on their interactions with other users in a community before selecting them for a leadership role.

Post	Mod	Timely\ Untimely	Unfriendly\ Friendly	Hostile\ Supportive	Unkind\ Kind	Guarded\ Open	Meant to Discuss?	Total	Notes
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
[NAME]			0	0	0	0	0	00	

Fig. 4. Moderator evaluation contingency scale (MECS).

This study did not evaluate private messages between moderators and users on Reddit because there is no access to these interactions. It is possible that a moderator engages a user differently in private message than he or she does on a public forum. This is a limitation of this study.

6   Conclusion

The Moderator Equivalency Contingency Scale, or MECS, is a strong indicator for evaluating moderation behavior, and it is a reliable instrument for measuring a moderator's ability to interact with users on a site. It measured with accuracy the ability for a moderator to positively interact with a user regarding the removal of content. This instrument has the potential to streamline evaluation of moderators for researchers as well as site administrators.

Different sites have different needs, and this study has only evaluated publicly viewable content in forums. As Chen, Xu, and Whinston found, moderator interactions may affect the type, quality, and frequency of content created [7]. The MECS can be used to evaluate potential moderators' interactions with other users to determine if their interaction style matches the needs of the site or forum as interaction style may vary minimally. Site administrators can also use MECS evaluations to provide feedback as to why a potential moderator may not have gained a coveted position.

References

1. Anders, G.: Technology- what is reddit worth?-The Web's antibullying, meme-spawning, porn-flecked link forum may go down as one of the shrewdest buys in the media business. Forbes, 56 (2012)
2. Avolio, B.J., Kahai, S.S.: Adding the "E" to E-Leadership: how it may impact your leadership. Organ. Dyn. **31**(4), 325–338 (2003). https://doi.org/10.1016/S0090-2616(02)00133-X
3. Avolio, B.J., Kahai, S., Dodge, G.E.: E-leadership: implications for theory, research, and practice. Lead. Q. **11**(4), 615–668 (2000)
4. Avolio, B.J., Sosik, J.J., Kahai, S.S., Baker, B.: E-leadership: re-examining transformations in leadership source and transmission. Lead. Q. **25**(1), 105–131 (2014)
5. Balthazard, P.A., Waldman, D.A., Warren, J.E.: Predictors of the emergence of transformational leadership in virtual decision teams. Lead. Q. **20**(5), 651–663 (2009)
6. Cascio, W.F., Shurygailo, S.: E-leadership and virtual teams. Organ. Dyn. (2003)
7. Chen, J., Xu, H., Whinston, A.B.: Moderated online communities and quality of user-generated content. J. Manag. Inf. Syst. **28**(2), 237–268 (2011)
8. Cherry, K.: Lewin's leadership styles. Psychology (2011). http://psychology.about.com/od/leadership/a/leadstyles.htm
9. Cheshire, C.: Online trust, trustworthiness, or assurance? Daedalus **140**(4), 49–58 (2011)
10. Cosley, D., Frankowski, D., Kiesler, S., Terveen, L., Riedl, J.: How oversight improves member-maintained communities. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 11–20 (2005)
11. Crosby, B.C., Bryson, J.M.: Integrative leadership and the creation and maintenance of cross-sector collaborations. Lead. Q. **21**(2), 211–230 (2010)
12. Day, D.V., Fleenor, J.W., Atwater, L.E., Sturm, R.E., McKee, R.A.: Advances in leader and leadership development: a review of 25 years of research and theory. Lead. Q. **25**(1), 63–82 (2014)
13. Delort, J., Arunasalam, B., Paris, C.: Automatic moderation of online discussion sites. Int. J. Electron. Commer. **15**(3), 9–30 (2011)
14. Dibbell, J.: Mutilated furries, flying phalluses: put the blame on griefers, the sociopaths of the virtual world. Wired Mag. **16**(2), 16–20 (2008)
15. Dinh, J.E., Lord, R.G., Gardner, W.L., Meuser, J.D., Liden, R.C., Hu, J.: Leadership theory and research in the new millennium: current theoretical trends and changing perspectives. Lead. Q. **25**(1), 36–62 (2014)
16. Ellison, N.B., Steinfield, C., Lampe, C.: The benefits of Facebook "friends:" social capital and college students' use of online social network sites. J. Comput. Mediat. Commun. **12**(4), 1143–1168 (2007)
17. Fiedler, F.E.: Predicting the effects of leadership training and experience from the contingency model. J. Appl. Psychol. **56**(2), 114 (1972)
18. Fiedler, F.E., Mahar, L.: A field experiment validating contingency model leadership training. J. Appl. Psychol. **64**(3), 247 (1979)
19. Gairín-Sallán, J., Rodríguez-Gómez, D., Armengol-Asparó, C.: Who exactly is the moderator? A consideration of online knowledge management network moderation in educational organisations. Comput. Educ. **55**(1), 304–312 (2010)
20. Gazan, R.: Understanding the rogue user. In: Nahl, D., Bilal, D. (eds.) Information and Emotion: The Emergent Affective Paradigm in Information Behavior Research and Theory. Information Today, pp. 177–185 (2007)

21. Gleave, E., Welser, H.T., Lento, T.M., Smith, M.A.: A conceptual and operational definition of "social role" in online community. In: 42nd Hawaii International Conference on System Sciences, pp. 1–11 (2009). https://doi.org/10.1109/HICSS.2009. 6
22. Grimes, J.M., Jaeger, P.T., Fleischmann, K.R.: Obfuscatocracy: a stakeholder analysis of governing documents for virtual worlds. First Monday **13**(9) (2008)
23. Huberman, B.A., Romero, D.M., Wu, F.: Social networks that matter: Twitter under the microscope. arXiv preprint arXiv:0812.1045 (2008)
24. Hussain, S.T., Lei, S., Akram, T., Haider, M.J., Hussain, S.H., Ali, M.: Kurt Lewin's change model: a critical review of the role of leadership and employee involvement in organizational change. J. Innov. Knowl. **3**(3), 123–127 (2018)
25. Huang, R., Kahai, S., Jestice, R.: The contingent effects of leadership on team collaboration in virtual teams. Comput. Hum. Behav. **26**(5), 1098–1110 (2010)
26. Kerr, S., Harlan, A.: Predicting the effects of leadership training and experience from the contingency model: some remaining problems. J. Appl. Psychol. **57**(2), 114–117 (1973)
27. Kirman, B., Lineham, C., Lawson, S.: Exploring mischief and mayhem in social computing or: how we learned to stop worrying and love the trolls. In: CHI 2012 Extended Abstracts on Human Factors in Computing Systems, pp. 121–130 (2012)
28. Lurey, J.S., Raisinghani, M.S.: An empirical study of best practices in virtual teams. Inf. Manag. **38**(8), 523–544 (2001)
29. Maslennikova, L.: Leader-centered versus follower-centered leadership styles. Lead. Adv. Online **14** (2007)
30. Mitchell, A.: Interventions for effectively leading in a virtual setting. Bus. Horiz. **55**(5), 431–439 (2012). https://doi.org/10.1016/j.bushor.2012.03.007
31. Purvanova, R.K., Bono, J.E.: Transformational leadership in context: face-to-face and virtual teams. Lead. Q. **20**(3), 343–357 (2009)
32. Reddit: Reddit Content Policy (2020). https://www.redditinc.com/policies/ content-policy
33. Tseng, J.J., Tsai, Y.H., Chao, R.C.: Enhancing L2 interaction in avatar-based virtual worlds: student teachers' perceptions. Australas. J. Educ. Technol. **29**(3), 357–371 (2013)
34. Vlachopoulos, P., Cowan, J.: Reconceptualising moderation in asynchronous online discussions using grounded theory. Distance Educ. **31**(1), 23–36 (2010)
35. Zaccaro, S.J., Bader, P.: E-leadership and the challenges of leading e-teams: minimizing the bad and maximizing the good. Organ. Dyn. **31**(4), 377–387 (2003)