Proceedings of the 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design

Multi-user Efficacy of Collaborative Virtual Environments

Aida Erfanian
Dept. of Electrical and Computer Engineering
University of Calgary
Calgary, Alberta, CANADA
aerfania@ucalgary.ca

Yaoping Hu
Dept. of Electrical and Computer Engineering
University of Calgary
Calgary, Alberta, CANADA
huy@ucalgary.ca

***Abstract*—Multi-user efficacy is a key factor of genuine collaboration among multiple users towards a common goal. To assess multi-user efficacy, social scientists have traditionally applied subjective measurements from a theoretical perspective. Researchers in human-computer interaction have developed combined metrics of objective and subjective measurements. Nevertheless, the combined metrics fall short to fully cover the theoretical perspective of social scientists. To remedy this shortfall, we have developed a set of objective and subjective metrics to complete the theoretical perspective. Utilizing the metrics, we present in this paper a study to verify the robustness of our dynamic priority (DP) model, which under a quasi-practical scenario resolves command conflicts and promotes perceived equality in interaction among multiple users. In the study, we utilized a realistic scenario which differs from the quasi-practical scenario in the allowance of verbal communication among users. The results of the study revealed that the DP model yielded a significantly higher degree of multi-user efficacy under the realistic scenario than the quasi-practical scenario. Moreover, there was no significant difference of the perceived equality in interaction between both scenarios. These observations confirm the robustness of the DP model, and imply the potential application of the model for genuine collaboration within multi-user VEs.**

Keywords- Multi-user efficacy; collaborative virtual environments; interaction models; verbal communication

I. INTRODUCTION

Computer-based virtual environments (VEs) are promising to promote genuine collaboration among multiple users of a working group. Underlying such collaboration are critical factors including awareness, usefulness, satisfaction, and efficacy [1]–[5]. Awareness among the users establishes a foundation of their collaboration; usefulness of a VE indicates its applicability of aiding the users to achieve their goals; satisfaction of the users highlights the degrees of fulfilling their needs; and efficacy of the VE signifies its ability of yielding desired collaboration. Together, these factors determine the usability of the VE for genuine collaboration.

Targeting awareness, research activities have attempted projecting the avatars of distributed users to be co-located together [6]. This projection allows all users to be proximate like in a co-located setting for their mutual cognizance. For usefulness, recent efforts have shifted to resolve command conflicts in multi-user interaction [2]. The conflict resolution relies apparently on interaction models to affect the usefulness. On satisfaction, a few studies have currently reported observations about the effect of perceived equality in interaction [7], [8]. It seems that this perception is a vital cognitive need. Regarding efficacy, there has been a vacuum of investigation however. This vacuum is left by the lack of a validated metrics for assessing multi-user efficacy.

Traditionally, efficacy has been a focus of scientists in social sciences [9]–[12]. Based on social cognitive theory [13], the scientists have classified individual and multi-user efficacy. Individual efficacy describes the capability of a user to undertake activities for achieving a goal [14]. Multi-user efficacy abstracts collective responses of multiple users to attain a common goal [12]. For assessing multi-user efficacy, the scientists have mainly utilized questionnaires to collect ratings perceived by the users on metrics – such as task performance, task focus, consensus, and workload [9]–[12]. That is, these metrics are subjective measurements.

In contrast, researchers in human-computer interaction (HCI) have the advantage of logging data using computers. This advantage allows the use of objective and subjective measurements to evaluate multi-user efficacy [4], [5], [15]–[17]. Based on logged data of user behaviors, the objective measurements are task performance [4], [5], [15], [16], task effectiveness and consensus [5][16]. Using questionnaires, the subjective measurements include the ease to perform tasks [17]. Except consensus, all measurements evaluate behaviors and responses of each individual user rather than collective behaviors and responses of multiple users. Although objective measurements complement subjective measurements by alleviating the reliance of user perception, the metrics used by HCI researchers fall short to fully cover the theoretical perspective of social scientists.

To mitigate the shortfall, we have developed a comprehensive set of metrics for assessing multi-user efficacy. The metrics consists of both objective and subjective measurements to complete the theoretical perspective. Differing from the metrics used by HCI researchers, each

The sponsor of this work is the Natural Sciences and Engineering Research Council of Canada (Grant No.: RGPIN-2015-06601 and STPGP 478870 – 15 to Y. Hu).

978-1-5090-1915-1/16/$31.00 ©2016 IEEE

TABLE 1. METRICS OF MULTI-USER EFFICACY.

	Measurements	Metrics	
Multi-user efficacy	Objective	Task Performance	Task completion time
			Ease to perform a task
		Task Effectiveness	Completeness
			Accuracy
		Consensus	
	Subjective	Perceived task focus	
		Perceived collaboration	
		Perceived workload	

metrics of our set measures collective behaviors and responses of multiple users. Moreover, we have conducted a human study to validate the set of metrics within a collaborative VE involving multiple users. In the study, we examined multi-user efficacy of the VE under a realistic scenario, compared to a quasi-practical scenario [2], [3]. Governed by the interaction model of dynamic priority (DP) [18] to resolve command conflicts, both scenarios emulated the collaboration of several domain experts (i.e., users) to achieve a common goal by working on a shared object. However, the realistic scenario differed from its counterpart in the allowance of verbal communication. The results of the study revealed that the realistic scenario yielded indeed higher multi-user efficacy than the quasi-practical scenario, due to verbal communication. The revelation validated our metrics of multi-user efficacy.

II. METRICS OF MULTI-USER EFFICACY

Consolidating both perspectives of social scientists and HCI researchers, we developed a comprehensive set of metrics for assessing multi-user efficacy. The metrics had objective and subjective measurements. Deriving from computer-logged data, the objective measurements included metrics of task performance, task effectiveness and consensus. Using questionnaire-based data, the subjective measurements were metrics of perceived task focus, collaboration and workload. Each of these metrics was dependent on collective behaviors and responses of all N users, who collaborate to achieve a common goal. Table 1 gives the metrics of multi-user efficacy.

For objective measurements, we defined task performance, *TP*, as a ratio ($TP = EPT/TCT$) of two sub-metrics: task completion time (*TCT*) and ease to perform a task (*EPT*). *TCT* is the average time taken by all N users of a working group to complete a designated task. *EPT* is the degree to which these users can accomplish collectively a task with minimum effort. This definition differs from its subjective counterpart for an individual user [17]. A higher *TP* denotes easier to perform a task, shorter completion time or both; leading to a more excellent task performance. We described task effectiveness, *Efftv*, as an amplification ($Efftv = Cmpl \times Acc$) of the following two sub-metrics: completeness (*Cmpl*) and accuracy (*Acc*). Completeness means a degree to which the users of the group have achieved their common goal; and accuracy indicates a level to which they have precisely carried out their designated tasks [3]. Thus, better task effectiveness is related to higher task completeness and accuracy. We represented the consensus, ξ, as a percentage of similarity in task behaviors among all users [2]. A larger value of ξ means a higher degree of consensus.

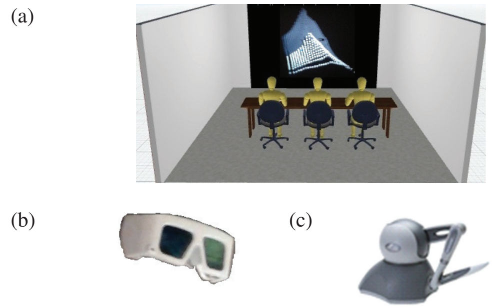

Fig. 1. The setup of the collaborative VE: (a) an example of multiple users working on a shared geological object in the display; (b) a pair of active stereo goggles; and (c) a haptic device of PHANToM Omni.

For subjective measurements, we employed three metrics of task focus, collaboration, and workload based on the users' perception. We converted ratings on two percentage bars of a questionnaire, representing respectively the perception of task focus and collaboration, to numeric values for computing their averages among the users. A larger value of the average means a higher degree of task focus or collaboration for the group. To calculate workload, we applied the NASA TLX [19]. A lower score of workload contributes to higher efficacy for the group.

III. STUDY METHODS

We conducted a human study to validate the metrics in Table 1, under both realistic and quasi-practical scenarios within a multi-user collaborative VE.

A. Collaborative Virtual Environment

As depicted in Fig. 1(a), a VE was developed for multiple users to simultaneously interact with a shared object. Using C++, the VE utilized a multi-threaded software architecture including: a management thread, a graphic thread and a haptic thread. The management thread coordinated both graphic and haptic threads in OpenGL and OpenHaptics libraries, respectively. The graphic thread rendered a virtual shared object on a wall-sized (3m x 3m), three-dimensional (3D) stereoscopic display. The shared object was a geological terrain with numerous cells. Each cell contained several geological properties of the terrain, which were represented by colors. Wearing a pair of active 3D stereo goggles as depicted in Fig. 1(b), each user saw the shared object in 3D stereoscopic view. The haptic thread handled the interactive commands issued by three users using three identical haptic devices, due to the availability of the devices. As shown in Fig.1(c), each device is a PHANToM Omni (SensAble Technologies Inc., USA). Using an Omni device, each user could issue interactive commands to interact with the shared object. For resolving command conflicts, we utilized the DP model [18] to warrant the users' perceived equal opportunities of interaction [3]. The VE ran on a computer with 2.53GHz processors (Intel© Xeon© dual quad core), a Quadro FX400 NVidia® graphics card, 4GB RAM, and a Windows® 7 operating system.

B. Participants

Using a within-subject design, the study had a total of 36 human participants (21 males and 15 females). On average, the participants were 26.47 ± 6.56 years old and naïve to the purpose of the study. The participants formed randomly 12 working groups (i.e., 3 participants per group), much more than needed 8 groups according to the sample size using the Lehr's formula. Within the VE, the 3 participants of each group functioned as 3 pseudo-experts to interact collaboratively with the shared object. All participants were right-handed determined by a modified version of Edinburgh handedness inventory. According to the Ishihara color blindness and the Randot Stereo test, none of the participants was color blinded; and they all had a normal or a corrected to normal vison with a stereo acuity of at least 40" arc. The study followed the Canadian Tri-council Ethics Guidelines.

C. Scenarios

A multi-user collaboration was conducted under a quasi-practical scenario [2], [3] and a realistic scenario. Both scenarios imitated a working meeting of three domain experts in the oil/gas industry. The experts are a geologist, a reservoir engineer and a production engineer. In practice, each expert undergoes different tasks towards a common goal. The term of "quasi-practical" referred to two aspects, which set the scenario departed from the practice. One aspect was the imitation of a domain expert by one participant as a pseudo-expert. Another aspect was the prohibition of verbal communication among the participants. The common goal of the pseudo-experts was to color collaboratively a map of geological properties on the shared object. To achieve the goal, each pseudo-expert had a unique list of designated tasks. That is, the list of each pseudo-expert represented a distinct knowledge of the property map. The three lists of the pseudo-experts were complementary to form the map. The completion of all designated tasks indicated 100% achievement of the goal. Thus, all pseudo-experts of a group were required to perform the tasks according to their lists actively and to observe each other's actions passively. Without verbal intervention, their collaboration departed from an actual collaborative meeting in practice.

In contrast, the realistic scenario permitted the pseudo-experts to communicate verbally to aid their collaboration, while they performed the same tasks as under the quasi-practical scenario. We designed a protocol of verbal communication for the pseudo-experts to follow. The protocol permitted the pseudo-experts to give guiding assistance to each other for completing a task. The protocol restricted the vocabularies in use and the ways of speaking them. The restrictions served for the following purposes: to vary controllably the levels of verbal affirmativeness; and to ensure the clear heard of the vocabularies. However, the pseudo-expert in action decided whether or not to follow a verbal guidance. The vocabularies were "yes" and "no" to characterize an affirmative mode and a counteractive mode, respectively. In one syllable, these vocabularies have distinctive vowel sounds to be spoken clearly and understood easily; and cover the whole spectrum of the affirmativeness from a positive to a negative. The way of speaking the vocabularies was a repetition as "yes yes yes" and "no no no". Such repetition is naturally to pronounce and easily to get attention.

While one pseudo-expert was acting on the shared object to complete a task, his/her peers could engage in one of the following three forms of assistance: (a) to use the affirmative mode for encouraging the pseudo-expert to maintain his/her course of actions; (b) to employ the counteractive mode for dissuading the pseudo-expert to continue his/her course of actions and for making a directional change; and (c) to keep silence without speaking any vocabularies for indicating a neutrality (or uncertainness) about the course of actions. However, the pseudo-expert decided to accept or discard the assistances. Hence, all pseudo-experts of the group undertook their collaboration by enacting their tasks, observing each other's actions and giving verbal guidance to assist task actions, if necessary. Thus, the realistic scenario abstracted a working meeting in practice better than the quasi-practical scenario.

D. Procedure

The procedure of the study had two sessions: one practice session prior to one testing session. During the practice session, all participants of the group learnt how to perform as pseudo-experts under the realistic scenario. They were instructed the vocabularies to use and the way to speak them. The practice session had 4-5 minute blocks, designed to teach and evaluate the participants against four criteria. The first criterion ensured that all participants of the group could recognize the occurrence of an interaction opportunity. This occurrence was signaled by a light turned from yellow to green on the display of the VE, and was completed by 3 logged pressings of the dark gray button on their Omni devices (i.e., interactive commands) within 485 ms of the green light. The DP model resolved the command conflicts by considering historical interactions of each participant to determine who gained the access to the shared object [18]. The second criterion evaluated each participant's skills in interacting with the object by using his/her Omni devices to complete a designated task in about 45 s. The participant who gained access to the shared object acted to rotate the object, to locate a given cell on the object and to color the cell. The third criterion assessed each participant's engagement in collaboration without verbal communication. Each participant needed to complete at least 4 tasks on his/her list within a block and to report his/her observations of the peers' actions on the display. The fourth criterion tested each participant to follow the verbal protocol for assisting the participant in action. There was at least one verbal assistance for over 50% of the interaction opportunities in a block. The pass of all criteria qualified the participants of the group to proceed to a questionnaire and then to the testing session.

The testing session was composed of two blocks, corresponding to the realistic scenario (DP_V) and the quasi-practical scenario (DP_NV). The order of the testing blocks was counterbalanced among all 12 groups. There was a dummy block between the two testing blocks, if the first block was the DP_NV block. Under the realistic scenario, the dummy block eliminated the side-effect of the non-verbal allowance on the next block. The time length of each block, including the dummy block, was fixed for 5 minutes long. The number of the completed tasks varied in each testing block, depending on how the participants completed their tasks. The common goal of each group was to accomplish a maximum of 21 tasks per block, equivalent to 7 tasks per participant. This task number balanced the testing length for each group and the complexity of the tasks of each expert in a working meeting. This complied with the observation reported in literature [3]. An identical questionnaire was administered to each participant after each testing block, including the dummy block.

Each block of the practice and testing sessions had a unique set of three task lists, corresponding to three domain experts in a working meeting. Between two blocks, there was a 3-minute break after completing the questionnaire. Each group spent totally 1.5 hours in the study, covering the pre-tests, the practice and testing sessions and the breaks.

E. Data Collection and Analysis

We used two methods of data collection – application logs and questionnaires. The VE application logs recorded data for computing the metrics of objective measurements in Table 1. For each group, the application logged all interaction opportunities, the pressing of the gray button by each participant at an interaction opportunity, the participant who gained access to the shared object at each interaction opportunity, task completion time of the participant, the action course of the participant on the shared object and tasks completed by each participant. The questionnaires gathered the perceived ratings of all participants during each testing block. These ratings were used to calculate the metrics of subjective measurements in Table 1. On the questionnaires, each participant answered a question by marking a vertical line on a bar bounded from 0% to 100%. There were three questions related to: a percentage of his/her gained interactions, a percentage of his/her task focus and quality of group collaboration. The vertical lines on each questionnaire were quantified for further analyses. For the perceived workload, the participant gave a score to each factor of the NASA TLX [19] for computing the workload.

Two baseline evaluations were conducted before analyzing each metrics of multi-user efficacy, as given in Table 1. The first baseline evaluation ensured that the logged data did not contain outliers (i.e., misbehavers' data). The absence of any outlier in a group implies that the total sum of all participants' percentages of gained interactions must be equal to 100%; that the mean of the percentages shall be 33.33% and that the standard deviation of the percentages should be zero under both DP_V and DP_NV blocks. From the logged data, we derived the percentages of interaction opportunities, at which each participant of the group gained access to the shared object. Then, we calculated the mean and standard deviation of the percentages for both DP_V and DP_NV blocks. Any outlier disqualified the data of the group for further analyses.

The second baseline evaluation verified whether the participants of all groups perceived an equality in interaction under both DP_V and DP_NV blocks. We used the "mean-std" technique reported in literature [2], [3], [18]. For each group of 3 participants, this technique requires to average the perceived percentages of interaction for all participants and to compute the standard deviation of the perceived percentages. If the average perceived percentages of interaction within a group are close to 33%, and the standard deviations of these percentages are close to zero; all participants indicate a perceived equality in interaction. Along with a normal distribution test, we used one-way analysis of variance (ANOVA) with repeated measures to verify that there was no significant difference of the perceived percentages between the testing blocks.

Once the baseline evaluations fulfilled their purposes, we conducted analyses on multi-user efficacy. The analyses

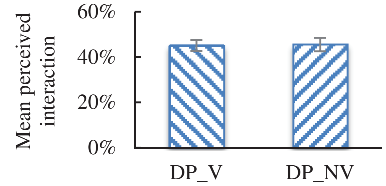

Fig. 2. Average percentages of gained interaction. [Error bars represent standard errors.]

computed all metrics of objective and subjective measurements, as described in Section II. Consequently, we carried out a one-way ANOVA with repeated measures to compare each metrics between the DP_V and DP_NV blocks. We conducted normality tests to ensure the data eligibility of each metrics for the ANOVA analyses.

IV. RESULTS

The first baseline evaluation revealed that no outlier existed among all groups for both DP_V and DP_NV blocks. The logged data of each group yielded a sum of 100% for the percentages of interactions; the mean of these percentages was 33.33% for all groups, with a standard deviation of zero. The second baseline evaluation confirmed that both blocks induced similar levels of the perceived equality in interaction. The normality test yielded a normal distribution of the perceived percentages of interaction. The ANOVA analysis denoted no significant difference of the perceived percentages between both blocks [F (1, 11) = 0.43; $p > 0.05$]. As shown in Fig. 2, the means of perceived percentages were 45 ± 2% and 46 ± 3% for the DP_V and DP_NV blocks, respectively.

Normality tests on the data of each sub-metrics and metrics in Table 1 confirmed the normal distribution of the data in both blocks. This laid the foundation of using one-way ANOVA analyses to compare data statistically.

For objective measurements, we compared the two sub-metrics of task completion time *TCT* and ease to perform a task *EPT* before comparing task performance *TP*. An ANOVA analysis on *TCT* indicated a significant difference between the DP_V and the DP_NV blocks [F (1, 11) = 6.73; $p < 0.05$]. Averagely, the *TCT* value of the DP_V and DP_NV blocks were 0.27 ± 0.02 minutes and 0.42 ± 0.06 minutes, respectively. An ANOVA analysis on *EPT* revealed a significant difference between both blocks [F (1, 11) = 25.88; $p < 0.001$]. The *EPT* value was on average 8 ± 1% of a task per action (denoted as task-action) for the DP_V block and 4 ± 1% task-action for the DP_NV block. That is, one action of each participant achieved about 8% of one task in the DP_V block. In contrast, the same action completed about 4% of one task in the DP_NV block. Applying ANOVA on *TP*, there was a significant difference between the blocks [F (1, 11) = 27.62; $p < 0.001$]. The DP_V and DP_NV blocks had *TP* values of 31 ± 4% and 13 ± 3% task-action per minute, respectively, as shown in Fig. 3.

Task effectiveness *Efftv* had also two sub-metrics of completeness *Cmpl* and accuracy *Acc*. An ANOVA analysis indicated a significant difference of *Cmpl* between the blocks [F (1, 11) = 25.26; $p < 0.001$], with the values of 88 ± 3% and 64 ± 6% for the DP_V and DP_NV blocks, respectively.

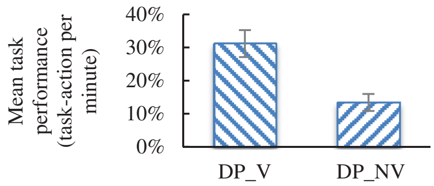

Fig. 3. Task performance of both DP_V and DP-NV blocks. [Error bars represent standard errors.]

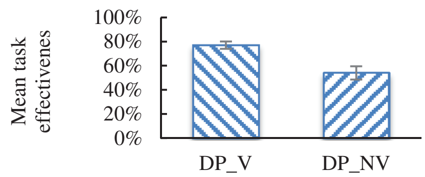

Fig. 4. Task effectiveness. [Error bars are standard errors.]

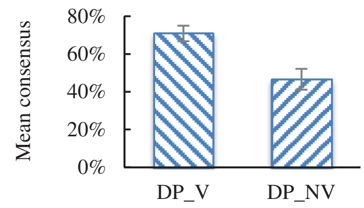

Fig. 5. Consensus. [Error bars represent standard errors.]

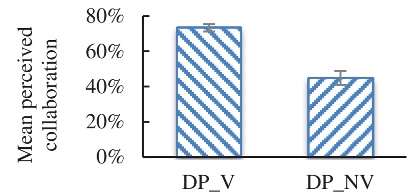

Fig. 6. Perceived collaboration. [Error bars are standard errors.]

However, a similar analysis on *Acc* did not reveal a significant difference between both blocks [F (1, 11) = 2.44; $p > 0.05$]. The DP_V block had its *Acc* value of 87 ± 2%, while the value of the DP_NV block was 84 ± 2%. Nevertheless, an ANOVA analysis on *Efftv* suggested a significant difference between both blocks [F (1, 11) = 34.17; $p < 0.001$]. As depicted in Fig. 4, the task effectiveness of the DP_V block was 77 ± 3%, compared to its counterpart of 54 ± 5% for the DP_NV block.

An ANOVA analysis gave a significant difference of consensus ξ between both blocks [F (1, 11) = 55.88; $p < 0.001$]. As shown in Fig. 5, the value of consensus was 71 ± 4% for the DP_V block and 47 ± 5% for the DP_NV block.

For subjective measurements, ANOVA analyses showed no significant difference of the perceived task focus between both blocks [F (1, 11) = 0.32; $p > 0.05$], with the values of 86 ± 3% and 87 ± 3% corresponding to the DP_V and DP_NV blocks. A similar result was found for the perceived workload [F (1, 11) = 3.46; $p > 0.05$]. The value of the perceived workload was 117.89 ± 14.89 for the DP_V block and 123.61 ± 16.49 for the DP_NV block. In contrast, there was a significant difference of the perceived collaboration between both blocks [F (1, 11) = 36.57; $p < 0.001$], with the value of 73 ± 2% for the DP_V block and 45 ± 4% for the DP_NV block, as depicted in Fig. 6.

V. Discussions

Table 2 summarizes the outcomes of all statistical analyses described above. Bolded "low" and "high" indicate the statistical significances of each pair-wise comparison. Four of all 6 defined metrics yielded significantly higher values in the DP_V block than in its counterparts of the DP_NV block. Three of these 4 distinguishable metrics are objective measurements, while the fourth metrics – perceived collaboration – is a subjective measurement. The rest two metrics of subjective measurements demonstrated an indistinguishability between both blocks. Hence, it is reasonable to assert that the DP_V block produced overall a higher level of multi-user efficacy than the DP_NV block. It is noteworthy that the higher level of the efficacy was derived from the verbal communication – the only difference between both blocks. This higher level of the efficacy is expected, because verbal communication is an intrinsic element of the human cognition to facilitate multi-user collaborations. The agreement between the observation of the study and the expectation validated our metrics of multi-user efficacy to assess collaborative VEs.

Moreover, our metrics remedy two deficiencies of the existing HCI metrics [4], [5], [15]-[17]. One deficiency is their consideration of individual user's behaviors and responses, except consensus. This deficiency would constrain the applicability of the HCI metrics to assess collective behaviors and responses of multiple users. Another deficiency is the lack of subjective measurements– albeit for easy to perform a task – to cover the perspective of social sciences. This perspective has traditionally been theorized to meet the human cognition [13]. Excluding the cognition would render the HCI metrics incapable of measuring multi-user synergies for collaboration. In contrast, each of our metrics are based on collective behaviors and responses of multiple users for both objective and subjective measurements. Evidently, the perceived collaboration among the users under the realistic scenario was 28% higher than under the quasi-practical scenario; even though the perceived task focus and workload were at similar levels under both scenarios. The collectiveness of subjective measurements was echoed by that of objective measurements as well. Hence, our metrics would be suitable for assessing cognitive synergies among multiple users. Although the study was conducted in a co-located setting, our metrics is free from this setting constraint. Thus, our set of metrics is applicable for assessing multi-user efficacy of both distributed and co-located collaboration.

Verbal and non-verbal communications are fundamental factors of human socialization. In turn, the factors might affect how an interaction model to manifest in the governance of multi-user interaction for collaboration. This was the rationale behind the second baseline evaluation. The outcomes of this evaluation revealed that verbal communication had no impact on the DP model to underlie the perceived equality in interaction. This revelation indicates a robustness of the DP model. Derived from the evaluation, it is noticed that both

TABLE 2. DP_V VS. DP_NV ON MULTI-USER EFFICACY.

<table>
<tr><th colspan="2">Metrics</th><th colspan="2">DP_V</th><th colspan="2">DP_NV</th></tr>
<tr><td rowspan="2">Task Performance</td><td>Task completion time</td><td>low</td><td rowspan="2">high</td><td>high</td><td rowspan="2">low</td></tr>
<tr><td>Ease to perform a task</td><td>high</td><td>low</td></tr>
<tr><td rowspan="2">Task Effectiveness</td><td>Completeness</td><td>high</td><td rowspan="2">high</td><td>low</td><td rowspan="2">moderate</td></tr>
<tr><td>Accuracy</td><td>high</td><td>high</td></tr>
<tr><td colspan="2">Consensus</td><td colspan="2">high</td><td colspan="2">low</td></tr>
<tr><td colspan="2">Perceived task focus</td><td colspan="2">high</td><td colspan="2">high</td></tr>
<tr><td colspan="2">Perceived collaboration</td><td colspan="2">high</td><td colspan="2">low</td></tr>
<tr><td colspan="2">Perceived workload</td><td colspan="2">moderate</td><td colspan="2">moderate</td></tr>
</table>

means and standard deviations departed slightly from their theoretical values. In an upward trend, the departure resulted largely from the overestimating tendency of the human perception and a relatively small number of the participant groups. On the basis of the DP model, the perceived equality in interaction under both realistic and quasi-practical scenarios agreed with the observations in literature under the quasi-practical scenario [2], [3]. In other words, the role of the DP model is independent of verbal communication. Yet, two aspects remains unknown. One aspect is how haptic feedback as a form of non-verbal communication could alter this robustness of the DP model; another aspect is how this feedback would affect the metric validation of multi-user efficacy. Future activities are to investigate these aspects.

VI. CONCLUSION

This work presented a novel set of metrics to assess multi-user efficacy of collaborative VEs. The metrics included objective and subjective measurements, and took account collective behaviors and responses of multiple users. A human study validated the metrics within a collaborative VE. These observations of the study imply a potential of using the metrics for assessing multi-user efficacy of collaborative VEs. Future work needs to examine how haptic feedback as non-verbal communication to affect the robustness of the DP model and multi-user efficacy.

REFERENCES

[1] S. Beck, A. Kunert, A. Kulik, et. al., “Immersive group-to-group telepresence,” IEEE Trans. Vis. Comput. Graphics, vol. 19, no. 4, pp. 616–25, Apr. 2013.

[2] A. Erfanian and Y. Hu, “Conflict resolution models on usefulness within multi-user collaborative virtual environments,” Proc. 10th IEEE 3DUI, Arles, France, 2015, pp. 147–148.

[3] A. Erfanian and Y. Hu, “The effect of interaction models on multi-user usability of collaborative virtual environments,” Proc. 11th IEEE CollaborateCom, Miami, FL, USA, 2014, pp. 187194.

[4] G. Wallis and J. Tichon, “Predicting the efficacy of simulator-based training using a perceptual judgment task versus questionnaire-based measures of presence,” Presence, vol. 22, no. 1, pp. 67–85, Feb. 2013.

[5] D. Powell and M. K. O’Malley, “The task-dependent efficacy of shared-control haptic guidance paradigms,” IEEE Trans. Haptics, vol. 5, no. 3, pp. 208–219, Aug., 2012.

[6] W. Chen, N. Ladeveze, D. Mestre, et. al., “User cohabitation in multi-stereoscopic immersive virtual environment for individual navigation tasks,” Proc. IEEE VR, Arles, France, 2015, pp. 47–54.

[7] Y. Rogers, Y. Lim, W. R. Hazlewood, et. al., “Equal opportunities: do shareable interfaces promote more group participation than single user displays?,” HCI, vol. 24, no. 1–2, pp. 79–116, Apr., 2009.

[8] L. Razmerita and K. Kirchner, “Social media collaboration in the classroom: a study of group collaboration,” Collaboration & Techno., vol. 8658 Lecture Notes Comput. Sci., pp. 279–286, 2014.

[9] D. I. Jung and J. J. Sosik, “group potency and collective efficacy: examining their predictive validity, level of analysis, and effects of performance feedback on future group performance,” Group & Organization Management, vol. 28, no. 3, pp. 366–391, Sept. 2003.

[10] D. W. S. Tai, Y.C. Hu, et. al., “The study on the influence model of professional teachers collective efficacy and student learning outcome in vocational senior high school,” Proc. 4th Int. Conf. CSSIT, Seoul, South Korea, 2009, pp. 325–329.

[11] S.W. Joe, Y.H. Tsai, C.P. Lin, et. al., “Modeling team performance and its determinants in high-tech industries: Future trends of virtual teaming,” Tech. Forecast. & Soc. Change, vol. 88, pp. 16–25, 2014.

[12] M. Salanova, A. M. Rodríguez-Sánchez, W. B. Schaufeli, et. al., “Flowing together: a longitudinal study of collective efficacy and collective flow among workgroups,” J. Psycho., vol. 148, no. 4, pp. 435–55, 2014.

[13] A. Bandura, “Social cognitive theory: an agentic perspective,” Ann. Rev. Psycho., vol. 52, pp. 1–26, 2001.

[14] M. E. Gist and T. B. Mitchell, “Self-efficacy: a theoretical analysis of its determinants and malleability,” Acad. Manag. Rev., vol. 17, no. 2, pp. 183–211, 1992.

[15] G. Griffin and R. Jacob, “Priming creativity through improvisation on an adaptive musical instrument,” Proc. 9th ACM Conf. Creativity & Cognition, Sydney, Australia 2013, p. 146-155.

[16] Y. Li, V. Patoglu, and M. K. O’Malley, “Negative efficacy of fixed gain error reducing shared control for training in virtual environments,” ACM Trans. Appl. Percep., vol. 6, no. 1, pp. 1–21, 2009.

[17] F. D. Davis, “Perceived usefulness, perceived ease of use, and user acceptance of information technology,” MIS Quart., vol. 13, no. 3, pp. 319–340, 1989.

[18] A. Erfanian, T. Zeng, and Y. Hu, “Dynamic strategies of conflict resolution on human perception of equality within multi-user collaborative virtual environments,” Proc. 9th IEEE CollaboarteCom, Austin, TX, USA, 2013, pp. 363–370.

[19] S. G. Hart, “Nasa-Task Load Index (NASA-TLX); 20 Years Later,” Proc. Hum. Factors & Ergono. vol. 50, pp. 904–908, Oct., 2006.