

Assessing teamwork in undergraduate education: a measurement tool to evaluate individual teamwork skills

Emily Britton, Natalie Simper*, Andrew Leger and Jenn Stephenson

Queen's University, Kingston, Canada

Effective teamwork skills are essential for success in an increasingly team-based workplace. However, research suggests that there is often confusion concerning how teamwork is measured and assessed, making it difficult to develop these skills in undergraduate curricula. The goal of the present study was to develop a sustainable tool for assessing individual teamwork skills, with the intention of refining and measuring these skills over time. The TeamUp rubric was selected as the preliminary standardised measure of teamwork and tested in a second year undergraduate course (Phase One). Although the tool displayed acceptable psychometric properties, there was concern that it was too lengthy, compromising student completion. This prompted refinement and modification leading to the development of the Team-Q, which was again tested in the same undergraduate course (Phase Two). The new tool had high internal consistency, as well as conceptual similarity to other measures of teamwork. Estimates of inter-rater reliability were within a satisfactory range, although it was determined that logistical issues limited the feasibility of external evaluations. Preliminary evidence suggests that teamwork skills improve over time when taught and assessed, providing support for the continued application of the Team-Q as a tool for developing teamwork skills in undergraduate education.

Keywords: teamwork; assessment; peer evaluation; self-evaluation; learning outcomes

Introduction

One of the main goals of post-secondary education is to prepare students for their chosen career path, equipping them with the requisite practical skills necessary for occupational success. Collaboration has always been an important aspect of professional work; however, as rapid technological advancement continues to shape society, and in particular our professional sector, teamwork and collaboration have become increasingly critical skills (Kozlowski and Bell 2003; Levy and Murnane 2012). Routine cognitive work and rudimentary manual labour are often relegated to computer programs and algorithms, while workers are now required to undertake more complex tasks requiring critical thinking and communication with others (Levy and Murnane 2012). Educational research has highlighted this shift towards an increasingly complex skillset, with a focus on delineating key twenty-first-century skills. Within these models, teamwork is consistently identified as a crucial component of success in today's professional sector (Dede 2010; Trilling and Fadel 2009). To emphasise this point, a survey conducted by the Association of American

*Corresponding author. Email: natalie.simper@queensu.ca

Colleges and Universities (AAC&U) found that 71% of employers identified 'teamwork skills and the ability to collaborate with others in diverse group settings' as a learning outcome that necessitated increased emphasis in college and university programmes (Hart Research Associates 2010, 2). Furthermore, the Ontario Ministry of Training Colleges and Universities (2014) has identified teamwork as an essential employability skill, highlighting its importance as a graduate attribute. Thus, it follows that the development of teamwork skills is an important outcome of undergraduate study, and a critical area of focus for educational research and development (Association of American Colleges and Universities 2014; Hughes and Jones 2011).

However, teamwork is difficult to quantify, as it must be inferred from myriad interrelated behaviours and attitudes. This complicates both instructors' and students' ability to track the development and performance of teamwork in the classroom. Additionally, inconsistency in the conceptualisation of teamwork further complicates the question of measurement and assessment (Brannick, Salas, and Prince 1997; Careau, Vincent, and Swaine 2014; Hughes and Jones 2011; Riebe et al. 2010). The term teamwork is conceptually vague, as it can be used in different contexts to refer to distinct things. For example, effective teamwork can refer to a successful product produced by a group of individuals working together to achieve a common goal (Riebe et al. 2010), while in other instances teamwork may refer to the nature and quality of individual contributions to a team effort, adopting a process-oriented focus (Hughes and Jones 2011). This ambiguity in what exactly constitutes high-quality teamwork, and how to identify and evaluate it in the classroom, obfuscates the instructor's ability to incorporate and assess teamwork as a learning outcome within their curriculum. As a result, educators often express a lack of confidence and willingness to teach and assess teamwork within their courses (Oliver 2011). A measurement tool that identifies performance criteria and behavioural markers indicative of excellence in teamwork is therefore highly valuable.

Few comprehensive tools currently exist for measuring and evaluating teamwork across diverse educational disciplines, further complicating the assessment process. Notwithstanding this shortage, one tool that is used is the Comprehensive Assessment of Team Member Effectiveness (CATME) (see: <http://www.catme.org>). The instrument is intended to clarify expectations for students, and provide meaningful feedback for students and instructors on team member performance, in order to reduce the confusion surrounding assessment (Ohland et al. 2012). Research by Hastie, Fahy, and Parratt (2014) examining the CATME has identified three key limitations of the tool: (1) the assessment process is controlled through the CATME website and thus the tool may not be altered in any way, (2) the items are ill-defined, impeding students' understanding of expectations, and (3) the ambiguous definitions may reduce validity by complicating interpretation.

In addition to the CATME, other tools have been developed with the goal of assessing collaboration among team members (Careau, Vincent, and Swaine 2014; Curran et al. 2011; Malec et al. 2007; Valentine, Nembhard, and Edmondson 2012). However, most of these tools were developed for application within a specific domain, most commonly for use in specialised medical settings. They are effective when applied in the contexts for which they are intended; yet due to the specificity of these environments there is limited generalisability to broad educational contexts. Moreover, most of these tools focus on teamwork as it relates to a working environment. Increasing specialisation in the knowledge and culture of specific professional groups often leads to challenges for effective inter-professional teamwork (Hall

2005). This highlights the importance of developing collaborative skills before entering the professional sector, so that graduates are prepared once immersed in these settings. Teaching students how to be effective team members at the post-secondary level equips them with skills that can then be enhanced and mastered within the workplace.

The goal of the current study was to develop a sustainable tool with acceptable psychometric indices of reliability and validity, to be used for developing and assessing teamwork in undergraduate courses. The overarching intention was to foster teamwork as a fundamental learning outcome of undergraduate education.

Method

The project was implemented in a second-year undergraduate Drama course at Queen's University in Eastern Ontario. This course was selected due to the instructor's interest and willingness to be involved in the project. Additionally, strong teamwork skills are a core learning outcome for graduates of the Queen's University Drama programme. The successful creation and execution of work in the theatre profession is predicated on effective, cohesive and high-functioning teams. For any given theatrical production, members typically work in smaller specialised 'teams' – acting ensembles, technical crews, creative team of designers, administrative and marketing teams – whose work must be seamlessly integrated into the larger product of the live performance. Furthermore, the instructor adopted a team-based learning approach to facilitate student learning and emphasise this outcome. The course was taught in a newly developed active learning classroom, with individual small group tables enabling ease of group interaction.

The study was granted ethical clearance according to the recommended principles of Canadian ethics guidelines and Queen's University policies.

Sample

The sample consisted of undergraduate students enrolled in a second-year drama course on theatre history and literature. There were 70 students enrolled in the course, 56 of whom were female and 14 of whom were male. The students were almost entirely of typical university age (18–21), however two students were older. The majority of students were in a Drama Major programme, as the course is a requirement for this degree plan.

At the beginning of the course, students were assigned to one of twelve teams, consisting of five to six students each. The instructor implemented a detailed process to form the teams in order to equalise and disperse various skills, traits and qualities between the groups. Sixty-five students consented to have their evaluations included as part of the study and 54 of these consented to be videotaped while engaging in a team-based activity. Five teams had all members consent to being recorded, totaling 24 students.

Phase One

Teaching teamwork

In week two of the semester, students engaged in a brainstorming exercise where they were instructed to create concept maps that represented their group definitions of what was necessary for effective teamwork. Concept maps, which are graphic

organisers that allow learners to communicate, relate and explore their understanding of a particular concept (Novak and Gowin 1984), have been demonstrated to be an effective means of collaborative learning (Novak and Cañas 2008). The purpose of this exercise was therefore to provide students with a focused opportunity to reflect on the specific skills and traits that contribute to a positive team experience. As a result, students were able to develop a collective sense of desirable behaviours and attitudes they would be responsible for as a team member.

The TeamUp rubric

There were few existing instruments that aligned with the project goals. However, the TeamUp rubric developed by Hastie, Fahy, and Parratt (2014), intended to assess individual teamwork skills among midwifery students, was identified as a promising tool for use in a broader undergraduate population. The TeamUp rubric was therefore selected as our preliminary standardised measure of teamwork.

The TeamUp rubric was developed from the AAC&U's Valid Assessment of Learning in Undergraduate Education (VALUE) teamwork rubric (Hastie, Fahy, and Parratt 2014). The rubric divides the construct of teamwork into five dimensions: *contributes to team project* (CTP), *facilitates the contributions of others* (FCO), *planning and management* (PM), *fosters a team climate* (FTC) and *manages potential conflict* (MC). Each dimension consists of seven items (35 items in total), articulated as a particular behavioural indicator or performance criteria. It was decided that select items (3, 5, 6 and 7) of the *manages potential conflict* section warranted a 'not applicable' option, as it is possible to have teams that experience no conflict; forcing students to choose a rating when the behaviour is simply not demonstrated can confound results.

Procedure

The course activities were primarily geared towards teamwork, but one specific collaborative task, carried out in week five of the semester, was selected for data collection. The course instructor, along with input from the research team, designed the task to be an authentic team-based activity where students were required to create a production proposal for a major theatre festival. Following this activity, the students completed their self- and peer evaluations and the course facilitators and independent research assistants completed the external evaluations. For reliability purposes, it was important that team members had worked together for a significant duration of time to allow for considerable interaction; the recommended minimum is four weeks (Strom and Strom 2011).

External evaluations were made based on the aforementioned half-hour-long collaborative task. Each of the five course facilitators was designated a team to monitor and used the TeamUp rubric to assess the teamwork skills demonstrated by each member. During this class time the task was also videotaped using camcorders, which were set up on tripods situated along the perimeter of the group in order to be as non-invasive as possible. The two research assistants used the TeamUp rubric to independently assess the teamwork skills displayed by each member in all five recordings. The research assistants watched the video once per team member, evaluating and rating only one individual at a time. Students were instructed to base their self- and peer evaluations on their own and their peers' teamwork skills

demonstrated over the period of the course. Evaluations were completed through an online learning management system to provide ease of completion, collection and data entry. Completion of these evaluations counted as a participation component of students' course professionalism grades. Moreover, the peer evaluations were aggregated, in order to de-identify ratings and protect anonymity, and subsequently returned to the students to provide formative feedback.

TeamUp results

The results presented are based on three sets of data: self-, peer and external evaluations. Supported by inter-rater reliabilities, the peer evaluations were calculated by averaging the ratings made by each of the team members. Similarly the external evaluations were calculated by averaging the facilitator's and two research assistants' ratings. For the external ratings, the limited duration of the task meant that the students didn't always have the opportunity to demonstrate certain criteria, and thus some items were not evaluated. In those cases, there were missing values in the data-set due to the assessment context.

Descriptive statistics and item discrimination

Item variance was calculated (see Table 1) to identify items with low variances that were unlikely to differentiate among individuals. The external evaluations had seven items with low variance ($<.30$), the peer evaluations had six, and the self-evaluations had none. Most items had mean response ratings near the midpoint of the scale. However, items FTC1 (*behaves politely*) and FTC3 (*treats team members respectfully*) displayed mean scores above 3.0 (at the high end of the scale) across all three data-sets, indicating that nearly all raters responded similarly to these items. Corrected item-total correlations were computed to identify items that did not differentiate between high and low scores on the measure overall. There were three low item-total correlations ($<.30$) for the external evaluations, two for the peer evaluations and two for the self-evaluations (see Table 2).

Factor structure

Factor analyses were performed to examine the underlying factor structure of the rubric. A principal components extraction method was used based on eigenvalues greater than one. All three data-sets yielded comparable results, indicating that the rubric had a consistent underlying factor structure. Examination of the scree plots indicated that the data levelled off after one factor and a majority of items loaded ($>.30$) onto the single factor, labelled 'teamwork.' There were a few anomalies within each data-set.

Reliability

Measures of internal consistency were computed using Cronbach's α . The scale yielded a coefficient of .94 using the self-reported ratings and .96 using the peer ratings, indicating very high internal consistency. It was not possible to calculate the internal consistency using external evaluations due to missing values in the data-set. Inter-rater reliability estimates were calculated using intra-class correlation

Table 1. TeamUp rubric item descriptive statistics.

Item	Self-evaluation		Peer evaluation		External evaluation	
	Mean (SD)	Variance	Mean (SD)	Variance	Mean (SD)	Variance
CTP1	2.92 (.74)	.54	2.74 (.62)	.38	2.36 (.48)	.23
CTP2	2.92 (.69)	.48	2.76 (.53)	.28	2.19 (.64)	.41
CTP3	2.28 (1.07)	1.14	2.40 (.65)	.42	1.75 (.86)	.74
CTP4	2.48 (.94)	.89	2.36 (.73)	.54	1.75 (.65)	.42
CTP5	2.97 (.80)	.63	2.91 (.61)	.38	2.28 (.66)	.43
CTP6	2.95 (.72)	.51	2.64 (.58)	.33	2.14 (.73)	.53
CTP7	2.33 (.98)	.96	2.26 (.63)	.39	1.82 (.62)	.39
FCO1	2.69 (.94)	.89	2.70 (.57)	.32	2.50 (.65)	.42
FCO2	2.67 (.98)	.96	2.71 (.62)	.39	2.49 (.63)	.40
FCO3	2.69 (.87)	.75	2.56 (.56)	.31	1.60 (.77)	.60
FCO4	2.92 (.74)	.55	2.80 (.61)	.38	1.78 (.51)	.26
FCO5	2.75 (.77)	.60	2.56 (.52)	.27	2.26 (.67)	.44
FCO6	2.67 (.80)	.63	2.45 (.51)	.26	1.74 (.79)	.63
FCO7	1.70 (1.09)	1.20	1.63 (.60)	.36	.72 (.82)	.66
PM1	2.62 (.96)	.92	2.31 (.60)	.36	1.49 (1.00)	1.00
PM2	3.05 (.68)	.46	2.84 (.58)	.33	2.24 (.70)	.49
PM3	3.02 (.73)	.53	2.89 (.64)	.40	2.53 (.72)	.51
PM4	1.80 (1.15)	1.32	1.78 (.84)	.71	1.36 (1.30)	1.68
PM5	2.30 (.96)	.91	2.00 (.64)	.41	1.72 (.91)	.82
PM6	2.21 (1.07)	1.14	1.94 (.73)	.53	1.07 (.71)	.50
PM7	2.59 (.92)	.85	2.51 (.52)	.27	1.67 (.82)	.68
FTC1	3.20 (.57)	.33	3.04 (.46)	.22	3.38 (.40)	.16
FTC2	2.78 (.99)	.99	2.63 (.59)	.34	1.00 (.37)	.14
FTC3	3.18 (.56)	.32	3.08 (.46)	.22	3.56 (.34)	.11
FTC4	2.79 (.88)	.77	2.69 (.59)	.35	2.46 (.43)	.18
FTC5	2.90 (.85)	.72	2.79 (.67)	.45	2.10 (.81)	.65
FTC6	2.20 (1.15)	1.32	2.17 (.64)	.41	.85 (.71)	.50
FTC7	2.92 (.86)	.74	2.88 (.64)	.41	2.38 (.58)	.34
MC1	2.85 (.79)	.63	2.54 (.62)	.38	2.17 (.77)	.60
MC2	2.65 (.95)	.91	2.55 (.57)	.33	1.90 (.40)	.16
MC3	2.18 (1.22)	1.49	2.06 (.79)	.63		
MC4	2.90 (.81)	.66	2.68 (.64)	.41	1.86 (1.00)	.99
MC5	2.30 (1.26)	1.59	2.01 (.86)	.75	.80 (.86)	.74
MC6	2.76 (.87)	.76	2.55 (.69)	.48	2.22 (.83)	.69
MC7	2.44 (1.15)	1.33	2.38 (.99)	.98		

Note: Missing values resulted from limited data for items that had a ‘not applicable’ option making the analysis invalid.

coefficients (ICC) measuring absolute agreement; reliability estimates are considered good when values are between .6 and .74, and excellent when values are between .75 and 1.0 (Hallgren 2012). The external rater ICC was .75, and more specifically the coefficient measuring agreement between the two research assistants was .91. The peer rater (4–5 peers per group) ICC was .76. Correlations between the classes of raters were also examined using mean scores. Self- and peer ratings were strongly correlated, $r(65) = .62, p < .01$, and peer and external ratings were moderately correlated $r(24) = .44, p < .05$.

Table 2. TeamUp rubric corrected item-total correlations.

Item	Self-evaluations	Peer evaluations	External evaluations
CTP1	.52	.72	.81
CTP2	.64	.77	.80
CTP3	.59	.65	.32
CTP4	.71	.79	.75
CTP5	.68	.78	.47
CTP6	.62	.72	.79
CTP7	.65	.79	.79
FCO1	.66	.85	.81
FCO2	.57	.54	-.21
FCO3	.70	.75	.79
FCO4	.60	.82	.68
FCO5	.61	.69	.84
FCO6	.67	.70	.71
FCO7	.52	.70	.61
PM1	.68	.62	.89
PM2	.50	.74	.70
PM3	.60	.79	.89
PM4	.76	.63	.38
PM5	.62	.79	.83
PM6	.56	.83	.74
PM7	.67	.64	.70
FTC1	.68	.67	.23
FTC2	.68	.63	.66
FTC3	.68	.72	.26
FTC4	.77	.72	.70
FTC5	.63	.81	.71
FTC6	.36	.78	.72
FTC7	.47	.68	.71
MC1	.49	.71	.88
MC2	.32	.79	.37
MC3	.27	.20	
MC4	.65	.82	.66
MC5	.39	.30	
MC6	.13	.49	
MC7	.31	.18	

Note: Missing values resulted from limited data for items that had a 'not applicable' option making the analysis invalid.

Pragmatic properties

In order for students to complete the self- and peer evaluations, they were required to rate themselves and each member of their team on all 35 items. Consequently, if a student was placed in a team with five peers this meant that they completed six separate evaluations multiplied by 35 items (210 ratings). Speculating that each student took approximately 15 s to make each rating, this would mean a completion time of approximately 50 min. Completing such a large volume of ratings may have resulted in loss of focus throughout the evaluation, and failure to consider each decision carefully and thoughtfully. Facilitators and research assistants, like the students, were required to make 210 independent ratings in order to complete the evaluations. Furthermore, research assistants made their ratings based on video recordings of the

half-hour-long task, watching the recording once per team member multiplied by six members, requiring approximately 3 h per team.

Qualitative analysis

The video recordings were also qualitatively coded using the TeamUp rubric as a framework, grouping specific behaviours as they related to the evaluation criteria. The findings from this content analysis revealed that certain behaviours and contributions were suggestive of multiple criteria. This recurring theme of multiple items being demonstrated by a single behavioural marker or contribution emphasised the need to refine the tool by combining and condensing similar items. Although a larger pool of items increases reliability, merging items reduced the amount of time required for completion, making the instrument more concise.

Refinement and modification of the TeamUp rubric

Despite displaying acceptable psychometric indices, the large number of items resulted in the TeamUp rubric being lengthy and burdensome to complete. In order for the rubric to be sustainable and practical for future use, the number of items needed to be reduced. This was critical so that raters considered responses carefully, and so the tool yielded manageable data for the instructor. Evaluation of the structure, items and qualitative coding informed this process. The following refinements were made:

- (1) *Synthesises the contributions of others* (FCO6) and *uses arguments and evidence in problem solving and decision-making* (PM7) were removed because they displayed minimal variance.
- (2) *Uses technologies effectively* (CTP3) was removed due to a low item-total correlation coefficient and minimal relevance based on qualitative analysis.
- (3) *Listens attentively to others without interrupting* (FCO2), *behaves politely* (FTC1) and *if necessary, asks team members to honour agreed expectations of behaviour* (MC3) were combined with similar items based on qualitative analysis.
- (4) The three items related to addressing conflict: *recognises and responds to suppressed or indirect conflict* (MC5), *manages conflict in ways that strengthen overall team cohesiveness and effectiveness* (MC6) and *addresses destructive conflict directly and constructively* (MC7), were combined because they did not load onto the single factor and were largely not applicable (see Tables 1 and 2).

Although there was no need for conflict management in this context, researchers determined that responding to and managing conflict was a necessary aspect of teamwork. However, the combination of the items was deemed appropriate in the interest of reducing instances where a rating of 'not applicable' was warranted, effectively reducing the frequency of missing values in the data. Additionally, the combination of these items was supported by the conceptual overlap observed in the content analysis.

Phase Two

The Team-Q

The modified tool was relabelled the Team-Q (see Table 3). The Team-Q is organised into five categories, which mimic the dimensions of the TeamUp rubric. Although the tool's factor structure indicated that teamwork consisted of only a single dimension, it was decided that organising the rubric into these categories enhanced clarity and readability. Each category consists of three items with the exception of the *fosters a team climate* category, which only has two, totaling fourteen items. The Team-Q, like the TeamUp rubric, frames teamwork as a process using positive behavioural language, where higher ratings indicate consistent demonstration of desirable teamwork skills. The use of behavioural descriptors provides students with explicit guidelines as to what the manifestation of particular teamwork skills looks like, reducing the possibility of ambiguous interpretation. Each item is rated on a five-point frequency scale, ranging from 1 (never) to 5 (always), labelled with a verbal frequency descriptor. (Note: it was decided that the scale range be changed from 0–4 to 1–5, for ease of data analysis; the numeric zero is not compatible with some statistical software). Use of the Team-Q is intended to:

- (1) Provide an evaluation of individual students' teamwork skill set,
- (2) Identify areas of strength and weakness,
- (3) Increase accountability among team members,
- (4) Allow instructors to assess student development of teamwork skills as a learning outcome of curriculum.

Procedure

The Team-Q was tested in the same second-year undergraduate course as Phase One. The procedure of the second implementation, completed in week 23 of the course, was identical to the first. Students again engaged in a half-hour-long video-recorded collaborative task, and self-, peer and external evaluations were collected; external evaluations targeted the half-hour task, while self- and peer ratings were based on behaviours and attitudes over the entirety of the course.

Team-Q results

Again results are based on three sets of data: self-, peer and external evaluations. Supported by inter-rater reliabilities, peer evaluations were calculated by averaging the ratings across all peers in a given team, while external evaluations were calculated by averaging the ratings of the facilitator and two research assistants.

Descriptive statistics and item discrimination

Evaluation of item variance (see Table 4) revealed that no items exhibited low variance (<.30) across all three data-sets, suggesting that all items discriminated among individuals displaying varying levels of the teamwork skills measured. All items yielded mean scores near the midpoints of the scale, with self-ratings ranging from 3.96 to 4.49, peer ratings from 3.75 to 4.27 and external ratings from 2.6 to 3.86.

Table 3. Team-Q survey. Survey prompt: Teamwork is behaviours under control of the individual team members. In your experience, how often does your peer demonstrate the following?

Component	Description	0 Never	1 Sometimes	2 Usually	3 Regularly	4 Always
Contributes to team project	Participates actively and accepts a fair share of the group work Works skillfully on assigned tasks and completes them on time Gives timely, constructive feedback to team members, in the appropriate format					
Facilitates contributions of others	Communicates actively and constructively Encourages all perspectives be considered and acknowledges contributions of others Constructively builds on contributions of others and integrates own work with work of others Takes on an appropriate role in group (e.g. leader, note taker)					
Planning and management	Clarifies goals and plans the project Reports to team on progress					
Fosters a team climate	Ensures consistency between words, tone, facial expression and body language Expresses positivity and optimism about team members and project					
Manages potential conflict	Displays appropriate assertiveness: neither dominating, submissive, nor passive aggressive Contributes appropriately to healthy debate Responds to and manages direct/indirect conflict constructively and effectively					

Table 4. Team-Q item descriptive statistics.

Item	Self-evaluation		Peer evaluation		External evaluation	
	Mean (SD)	Variance	Mean (SD)	Variance	Mean (SD)	Variance
CTP1	4.19 (.93)	.87	4.11 (.78)	.61	3.83 (.80)	.64
CTP2	4.39 (.84)	.71	4.32 (.71)	.51	3.81 (.80)	.64
CTP3	3.96 (.92)	.86	3.81 (.70)	.50	3.15 (.89)	.79
FCO1	4.44 (.78)	.61	4.26 (.72)	.52	3.87 (1.00)	1.01
FCO2	4.49 (.60)	.36	4.06 (.73)	.54	3.06 (.96)	.93
FCO3	4.37 (.75)	.56	4.16 (.65)	.43	3.33 (.85)	.73
PM1	4.28 (.86)	.74	4.22 (.73)	.53	3.52 (.94)	.89
PM2	4.28 (.73)	.53	3.89 (.75)	.56	3.38 (.96)	.91
PM3	3.98 (.95)	.91	3.76 (.73)	.53	3.01 (1.08)	1.17
FTC1	4.32 (.78)	.61	4.15 (.67)	.45	3.35 (.95)	.89
FTC2	4.33 (.95)	.91	4.11 (.74)	.54	3.39 (1.13)	1.27
MC1	4.33 (.69)	.48	4.01 (.77)	.59	3.17 (1.10)	1.20
MC2	4.39 (.90)	.81	4.09 (.80)	.63	3.13 (.98)	.96
MC3	4.40 (.86)	.75	4.27 (.81)	.66	2.61 (.77)	.59

Although no items had mean ratings at the extremes of the range, it is notable that both self-, $t(21)=7.1$, $p < .01$, and peer ratings, $t(21)=6.79$, $p = < .01$, were significantly higher than the external ratings. Corrected item-total correlations (see Table 5) were high across all items in all three data-sets, with the exception of item FCO2 (*encourages all perspectives be considered and acknowledges contributions of others*), which had a low item-total correlation ($<.30$) when examining the self-reported data.

Factor structure

Factor analyses were performed with all three data-sets, using a principal components analysis based on eigenvalues greater than one. An examination of the scree plots indicated that the data levelled off after a single factor, suggesting that a one-factor model was the best fit for the data across all three sets of ratings. This

Table 5. Team-Q corrected item-total correlations.

Item	Self-evaluations	Peer evaluations	External evaluations
CTP1	.78	.91	.88
CTP2	.70	.84	.82
CTP3	.61	.81	.57
FCO1	.73	.95	.58
FCO2	.22	.82	.82
FCO3	.75	.91	.52
PM1	.64	.88	.77
PM2	.54	.84	.68
PM3	.59	.85	.68
FTC1	.58	.82	.74
FTC2	.58	.86	.63
MC1	.55	.81	.78
MC2	.62	.91	.61
MC3	.75	.92	.61

provided further evidence that the teamwork construct consists of only one dimension. Furthermore, all items loaded onto the single factor ($>.30$) across all three data-sets, again with the exception of FCO2, which did not load onto the single factor when examining the self-reported ratings. The teamwork factor based on external ratings had an eigenvalue of 6.66 and accounted for 47.58% of the item variance, based on peer ratings an eigenvalue of 10.99 and 78.46% of the item variance, and based on self-ratings an eigenvalue of 7.79 and 55.63% of the item variance.

Reliability

The scale exhibited high internal consistency across all three data-sets. The scale Cronbach’s α was .91 using self-reported ratings, .97 using peer ratings and .93 using external ratings. Inter-rater reliability estimates were again within an acceptable range, with an external rater ICC of .76, and .77 when examining just the two research assistants. The peer rater (4–5 peers) ICC was .69. Self- and peer ratings were also strongly correlated, $r = .54, p < .01$, however the correlation between peer and external ratings was weak, $r = .13, p = .56$. See Figure 1 for a summary of the reliability statistics.

Validity

The results of the factor analyses provided evidence of discriminant validity. The one-factor model indicated that the tool had limited association with other factors (all items loaded highly on the single factor) and measured one distinct construct. Content validity was assessed by comparing the items of the Team-Q with the core teamwork dimensions identified in a meta-analysis of existing instruments for measuring teamwork in health care (see Table 6). After consultation with educational researchers, it was concluded that there was a substantial amount of overlap between the Team-Q items and the thirteen teamwork dimensions identified by Valentine, Nembhard, and Edmondson (2012), suggesting that the tool had conceptual similarity with other measures of teamwork.

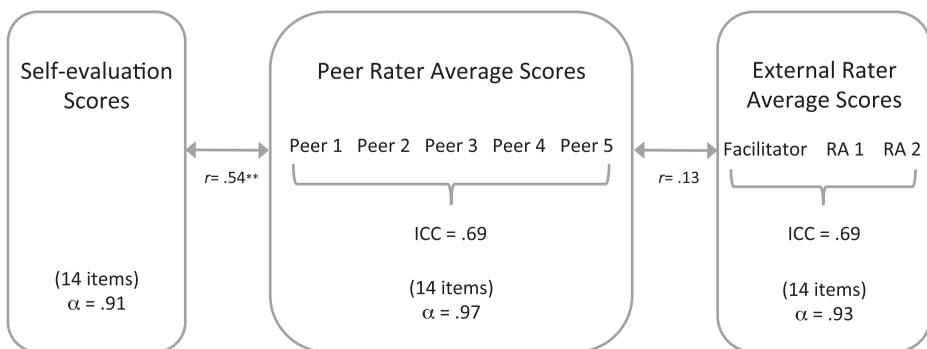


Figure 1. Overview of internal consistency and inter-rater reliability.

Pragmatic properties

The Team-Q, with its reduced number of items, took significantly less time to complete. In order to complete six evaluations (one self-evaluation and five peer evaluations) students were required to make 84 ratings, which based on previous calculations took approximately 20 min. The reduced time commitment meant that students were more likely to remain focused and engaged when completing their evaluations. Nevertheless, the research assistants were still required to watch the video recordings once per team member multiplied by six members, which remained a very lengthy process. This suggested that the tool might be more suitable for self- and peer evaluation moving forwards.

Improvement over time

In order to evaluate students' teamwork skills over the duration of the course, ratings from Phase One were converted to a scale ranging from 1 to 5 to be consistent with the range of the Team-Q. Student's mean scores across all items were then calculated and paired samples t-tests were conducted to compare student ratings from week 5 (Phase One) and week 23 (Phase Two). Figure 2 provides a summary of the results. There was a significant improvement in students' self-reported teamwork skills, $t(53) = 7.32, p < .01$, as well as a significant improvement in students' teamwork skills as reported by their peers, $t(63) = 8.43, p < .01$. There was a modest improvement when comparing the external evaluations from time one to time two, however the difference was not significant, $t(17) = 1.56, p = .14$.

Feedback from both students and the course instructor were consistent with these findings. Although student comments on the collaborative aspect of the course were not solicited, course evaluations received by the instructor following its completion overwhelmingly included references to teamwork. Many students expressed positivity about the teamwork aspect of the course, and felt that it facilitated their learning and performance. Furthermore, students felt that they were able to enhance key teamwork skills, such as collaborative idea generation, communication and facilitating positive interaction. For example, one student noted: 'I liked the group work and being in teams because I felt that it helped me expand my ideas and communication skills', while another student noted: '[I] loved being with my [team] mates, and had some great experiences working with them'. The instructor also provided positive feedback in regard to students' growth, expressing that students demonstrated increased meta-cognitive awareness of teamwork skills, which enhanced team behaviour. As such, the instructor felt that the use of the Team-Q proved to be beneficial, and expressed a desire to continue incorporating it into the course structure in future years.

Discussion

The cultivation of effective teamwork skills in post-secondary education prepares students to enter a workforce in which collaboration and team-based efforts have become ubiquitous. Employers and educators have identified the importance of developing these essential skills by incorporating teamwork as a learning outcome in curricula, however the teaching and assessment of teamwork in the classroom does not always support this outcome. Teamwork is a dynamic skill that can be

Table 6. Comparison of Team-Q items with teamwork dimensions identified in a meta-analysis by Valentine, Nembhard, and Edmondson (2012).

Team-Q items	Teamwork dimensions identified in meta-analysis by Valentine, Nembhard, and Edmondson (2012)											
	Communication	Shared decision-making	Use of all members' expertise	Full participation	Collaboration	Learning orientation	Coordination and effort	Social support	Respect (for members and their contributions)	Psychological safety	Active conflict management	Group cohesion
Participates actively and accepts a fair share of the group work				+		+	+					
Works skillfully on assigned tasks and completes them on time				+			+					
Gives timely, constructive feedback to team members, in the appropriate format	+				+			+		+		
Communicates actively and constructively	+				+							+
Encourages all perspectives be considered and acknowledges contributions of others		+	+									+
Constructively builds on contributions of others and integrates own work with work of others	+				+							+
Takes on an appropriate role in group (e.g. leader, note taker)						+						
Clarifies goals and plans the project		+		+								
Reports to team on progress	+	+		+								
	+											+

(Continued)

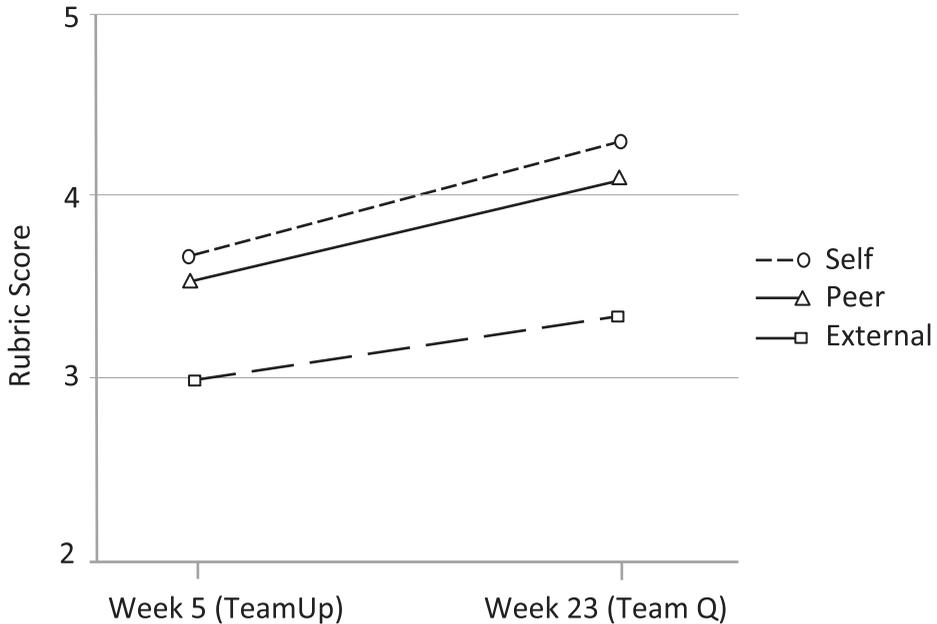


Figure 2. Student improvement in teamwork skills over time.

improved with continued application, evaluation and attention to shortcomings; over time students can learn to be more effective team members (Biggs and Tang 2011; Brutus and Donia 2010). The current study addressed the challenge of developing a practical and sustainable tool for assessing individual teamwork skills in order to enhance these skills as an outcome in undergraduate curricula.

The first step in this process was the implementation of the TeamUp rubric in an undergraduate drama course. Statistical analyses were performed on the data collected, which informed the modification and refinement of the rubric leading to the development of the Team-Q (see Table 3). The primary issue was the length of time required for completion of the evaluations, as the TeamUp rubric in its original form was lengthy and there was concern that this might compromise thoughtful completion. In developing the Team-Q the number of items was significantly reduced from 35 to 14 by removing problematic items and combining conceptually similar items.

In the second phase of the study, the newly developed Team-Q was implemented in the same undergraduate course. Factor analyses revealed that the teamwork construct consists of only one dimension, which was consistent with the initial analysis of the TeamUp rubric. Reliability analyses indicated very high internal consistency for the measure, providing evidence that the items were closely related. This was further confirmed by examining the instrument's discriminant validity through factor analysis and content validity, which both indicated that the Team-Q was conceptually distinct from other factors and conceptually related to other measures of teamwork.

Estimates of inter-rater reliability indicated that there was a satisfactory level of agreement across peer raters, as well as external raters. However, consistency was somewhat lower when comparing between different groups of raters, such that

comparison of self-peer and peer-external ratings yielded moderate-to-weak correlation coefficients respectively. The discrepancy between peer and external ratings likely stemmed from the peer ratings being based on behaviours over the duration of a full-year course (i.e. 23 weeks), while external evaluations were based on a single half-hour period. Peer ratings and self-ratings were likely a more accurate representation of individuals' teamwork skills based on this distinction; ratings over a longer temporal period tend to be more reliable (Rothstein 1990). Furthermore, the length of time required to complete the peer and self-evaluations was significantly shorter than the external evaluations. This suggested that the comprehensive assessment performed by the external raters was not practical or sustainable moving forwards. Therefore, future use of the tool is recommended for self- and peer evaluation.

Comparison of evaluations from Phase One (TeamUp) to Phase Two (Team-Q) revealed that students' average teamwork ratings displayed a significant increase over time, when examining both self- and peer evaluations. These results provide evidence to suggest that student's teamwork skills improved when teamwork was explicitly taught and assessed. It is notable that the elimination of problematic items between the first and second ratings may have influenced this, and it will be important to consider this in future evaluations. Furthermore, although improvement of teamwork skills was not a primary focus of the present study, comments from the instructor and students seem to indicate positive growth. The course instructor indicated that the explicit foregrounding of teamwork skills and exposure to the tool made students more mindful of what good teamwork looks like. Examination of students' course feedback further highlighted this point, as 64% of students' referenced teamwork in their open-ended responses even though they were not explicitly prompted to do so. Students who included such comments expressed enthusiasm about the teamwork component of the course, and awareness of its positive impact on their growth and learning.

Recommendations

Future research should consider limitations of the present study, namely a relatively small sample size drawn from a single undergraduate drama course. Diverse educational disciplines often exhibit a great deal of variation, with different course structures and types of assignments. For the purposes of this study, drama students provide an ideal cohort, being given to confident communication and group interaction. Team-based learning tasks like rehearsal of scenes, construction and installation of production elements like lighting and scenery, and conceptual collaboration in the creation of performances characterise course work typically undertaken by drama students. As a result, the students observed in this study may have been more engaged in the team process, while students in disciplines with a more empirical focus might initially place less value on team skills. The Team-Q is intended for use across a broad undergraduate population from diverse faculties and departments. Therefore, it is necessary to test the tool in a broad range of undergraduate courses to ensure that its properties are consistent across various educational contexts. As teamwork skills continue to be taught and assessed using the Team-Q in a wider sample of courses, further evaluation can confirm these findings.

Conclusions

The Team-Q was developed as a sustainable tool for evaluating individual teamwork skills, and to facilitate the development of these skills in undergraduate education. It was adapted from the TeamUp rubric developed by Hastie, Fahy, and Parratt (2014), with the number of items reduced to improve the practicality of use while retaining reliability. The tool displayed acceptable psychometric indices of reliability and validity, and is best suited for self- and peer evaluation. The explicit teaching of teamwork and use of the Team-Q provided students with clear guidelines about what constitutes effective teamwork, and performance evaluations to gauge students' skill development. The Team-Q holds promise for improving how teamwork is assessed in undergraduate education. It can help to align intended outcomes and assessment, and provide instructors with a practical method for assessing teamwork. It is assumed that development and mastery of teamwork skills will enhance student success by preparing graduates for the workplace.

Acknowledgements

We thank the anonymous reviewers whose supportive comments and suggestions helped improve an earlier draft of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Emily Britton is a research assistant with the Centre for Teaching and Learning, Queen's University, Kingston. She has been actively involved in both the Queen's Active Learning Project and Learning Outcomes and Assessment Project. Specifically, her current research focus is on the development and assessment of institutional learning outcomes and student teamwork.

Natalie Simper coordinates the Learning Outcomes and Assessment Project, investigating the development and measurement of meta-cognition and cognitive skills with learning outcomes. Natalie comes from an Australian senior-secondary/post-secondary teaching background, with experience at the statewide level in curriculum development, large-scale assessment and evaluation and assessment of outcome-based education.

Andrew Leger is an educational developer with the Centre for Teaching and Learning. He works closely with faculty and staff on the development of learning spaces, educational technologies and ways to enhance the student and educator experience through the implementation of these resources. Andy helps coordinate the use of the newly renovated active learning classrooms and is a member of the onQ Learning Management System Steering Committee.

Jenn Stephenson is an associate professor in the School of Drama and Music at Queen's University, where she teaches courses on dramatic literature, history and theory. Her book *Performing Autobiography: Contemporary Canadian Drama* (U of Toronto P, 2013) received the Ann Saddlemyer Award from CATR. Recent articles have appeared in *Theatre Journal*, *Theatre Research in Canada*, *New Theatre Quarterly* and *English Studies in Canada*.

References

- Association of American Colleges and Universities. 2014. "Essential Learning Outcomes." Association of American Colleges & Universities. Accessed November 25, 2014. <https://www.aacu.org/leap/essential-learning-outcomes>
- Biggs, J., and C. Tang. 2011. *Teaching for Quality Learning at University: What the Student Does*. Maidenhead: McGraw-Hill Education.
- Brannick, M. T., E. Salas, and C. W. Prince, eds. 1997. *Team Performance Assessment and Measurement: Theory, Methods, and Applications*. Mahwah, NJ: Taylor & Francis.
- Brutus, S., and M. B. Donia. 2010. "Improving the Effectiveness of Students in Groups with a Centralized Peer Evaluation System." *Academy of Management Learning & Education* 9 (4): 652–662.
- Careau, E., C. Vincent, and B. R. Swaine. 2014. "Observed Interprofessional Collaboration (OIPC) during Interdisciplinary Team Meetings: Development and Validation of a Tool in a Rehabilitation Setting." *Journal of Research in Interprofessional Practice and Education* 4 (1): 1–19.
- Curran, V., A. Hollett, L. M. Casimiro, P. McCarthy, V. Banfield, P. Hall, K. Lackie, I. Oandasan, B. Simmons, and S. Wagner. 2011. "Development and Validation of the Interprofessional Collaborator Assessment Rubric ((ICAR))." *Journal of Interprofessional Care* 25 (5): 339–344.
- Dede, C. 2010. "Comparing Frameworks for 21st Century Skills." In *21st Century Skills: Rethinking How Students Learn*, edited by James Bellanca and Ron Brandt, 51–76. Bloomington, IN: Solution Tree Press.
- Hall, P. 2005. "Interprofessional Teamwork: Professional Cultures as Barriers." *Journal of Interprofessional Care* 19 (S1): 188–196.
- Hallgren, K. A. 2012. "Computing Inter-rater Reliability for Observational Data: An Overview and Tutorial." *Tutorials in Quantitative Methods for Psychology* 8 (1): 23–34.
- Hart Research Associates. 2010. "Raising the Bar: Employers' View on College Learning in the Wake of the Economic Downturn." Association of American Colleges and Universities. Accessed June 15, 2015. http://www.aacu.org/leap/documents/2009_EmployerSurvey.pdf
- Hastie, C., K. Fahy, and J. Parratt. 2014. "The Development of a Rubric for Peer Assessment of Individual Teamwork Skills in Undergraduate Midwifery Students." *Women and Birth* 27 (3): 220–226.
- Hughes, R. L., and S. K. Jones. 2011. "Developing and Assessing College Student Teamwork Skills." *New Directions for Institutional Research* 2011 (149): 53–64.
- Kozlowski, S. W., and B. S. Bell. 2003. "Work Groups and Teams in Organizations." *Handbook of Psychology*. Accessed October 19, 2015. <http://digitalcommons.ilr.cornell.edu/articles/927>
- Levy, F., and R. J. Murnane. 2012. *The New Division of Labor: How Computers are Creating the Next Job Market*. Princeton, NJ: Princeton University Press.
- Malec, J. F., L. C. Torsher, W. F. Dunn, D. A. Wiegmann, J. J. Arnold, D. A. Brown, and V. Phatak. 2007. "The Mayo High Performance Teamwork Scale: Reliability and Validity for Evaluating Key Crew Resource Management Skills." *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare* 2 (1): 4–10.
- Novak, J. D., and A. J. Cañas. 2008. "The Theory Underlying Concept Maps and How to Construct and Use Them." Accessed October 19, 2015. <http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf>
- Novak, J. D., and D. B. Gowin. 1984. *Learning How to Learn*. Cambridge: Cambridge University Press.
- Ohland, M., M. Loughry, D. Woehr, C. Finelli, L. Bullard, R. Felder, R. Layton, H. Pomeranz, and D. Schmucker. 2012. "The Comprehensive Assessment of Team Member Effectiveness: Development of a Behaviorally Anchored Rating Scale for Self and Peer Evaluation." *Academy of Management Learning & Education* 11 (4): 609–630.
- Oliver, B. 2011. "Assuring Graduate Outcomes." *PowerPoint Presentation*. The Australian Learning and Teaching Council. Accessed August 27, 2015. http://www.public.jcu.edu.au/events/idc/groups/public/documents/conference_paper/jcu_086521.pdf

- Ontario Ministry of Training Colleges and Universities. 2014. "Essential Employability Skills." Ontario Ministry of Training Colleges and Universities. Accessed November 25, 2014. <http://www.tcu.gov.on.ca/pepg/audiences/colleges/progstan/essential.html>
- Riebe, L., D. Roepen, B. Santarelli, and G. Marchioro. 2010. "Teamwork: Effectively Teaching an Employability Skill." *Education + Training* 52 (6/7): 528–539.
- Rothstein, H. R. 1990. "Interrater Reliability of Job Performance Ratings: Growth to Asymptote Level with Increasing Opportunity to Observe." *Journal of Applied Psychology* 75 (3): 322–327.
- Strom, P. S., and R. D. Strom. 2011. "Teamwork Skills Assessment for Cooperative Learning." *Educational Research and Evaluation* 17 (4): 233–251.
- Trilling, B., and C. Fadel. 2009. *21st Century Skills: Learning for Life in Our times*. San Francisco, CA: Wiley.
- Valentine, M. A., I. M. Nembhard, and A. C. Edmondson. 2012. "Measuring Teamwork in Health Care Settings: A Review of Survey Instruments." *Medical Care* 53 (4): 16–30.

Copyright of *Assessment & Evaluation in Higher Education* is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.