Expert validation of a teamwork assessment rubric: a modified Delphi study

Jennifer A. Parratt
_Southern Cross University_

Kathleen M. Fahy
_Southern Cross University_

Marie Hutchinson
_Southern Cross University_

Gui Lohmann
_Southern Cross University_

Carolyn R. Hastie
_Southern Cross University_

See next page for additional authors

Publication details

Published version available from:
http://dx.doi.org/10.1016/j.nedt.2015.07.023
Expert Validation of a Teamwork Assessment Rubric: A Modified Delphi Study

Jenny A. Parratt, Adjunct Associate Professor a; Kathleen M. Fahy, Professor a; Marie Hutchinson, Associate Professor b; Gui Lohmann, Senior Lecturer c; Carolyn R. Hastie, Senior Lecturer a; Marilyn Chaseling, Associate Professor d; Kylie O’Brien, Director of Integrative Chinese Medicine, National Institute of Integrative Medicine e, Adjunct Associate Professor Victoria University, Fellow NICM University Western Sydney

Email: jennifer.parratt@scu.edu.au

a School of Health and Human Sciences, Southern Cross University, Locked Mail Bag 4 Coolangatta QLD 4225 AUSTRALIA
b School of Health and Human Sciences, Southern Cross University, PO Box 157, Lismore NSW 2480 AUSTRALIA
c School of Natural Sciences, Griffith University, 170 Kessels Road, Nathan, QLD 4111 AUSTRALIA
d Centre for Children and Young People, School of Education, Southern Cross University, PO Box 157, Lismore NSW 2480 AUSTRALIA
e National Institute of Integrative Medicine, 11-23 Burwood Rd Hawthorn, Victoria AUSTRALIA

NOTICE: this is the author’s version of a work that was accepted for publication in Nurse Education Today. Changes resulting from the publishing process, such as editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. A definitive version was subsequently published in Nurse Education Today.

The full reference is:

HIGHLIGHTS

- ‘TeamUP’ supports university students use of effective teamwork skills
- The theoretically grounded TeamUP Rubric defines and describes teamwork skills
- We use the TeamUP Rubric to teach and assess students doing team-based assignments
- Professionals with teamwork expertise have contributed to rubric validation
- The TeamUP Rubric is a well validated measure of tertiary student teamwork skills
ABSTRACT

Background: Teamwork is a ‘soft skill’ employability competence desired by employers. Poor teamwork skills in healthcare have an impact on adverse outcomes. Teamwork skills are rarely the focus of teaching and assessment in undergraduate courses. The TeamUP rubric is a tool used to teach and evaluate undergraduate students’ teamwork skills. Students also use the rubric to give anonymised peer feedback during team-based academic assignments. The rubric’s five domains focus on planning, environment, facilitation, conflict management and individual contribution; each domain is grounded in relevant theory. Students earn marks for their teamwork skills; validity of the assessment rubric is critical.

Question: To what extent do experts agree that the TeamUP rubric is a valid assessment of ‘teamwork skills’?

Design: Modified Delphi technique incorporating Feminist Collaborative Conversations.

Participants: A heterogeneous panel of 35 professionals with recognised expertise in communications and/or teamwork.

Method: Three Delphi rounds using a survey that included the rubric were conducted either face-to-face, by telephone or online. Quantitative analysis yielded item content-validity indices (I-CVI); minimum consensus was pre-set at 70%. An average of the I-CVI also yielded sub-scale (domain) (D-CVI/Ave) and scale content-validity indices (S-CVI/Ave). After each Delphi round qualitative data were analysed and interpreted; Feminist Collaborative Conversations by the research team aimed to clarify and confirm consensus about the wording of items on the rubric.

Results: Consensus (at 70%) was obtained for all but one behavioural descriptor of the rubric. We modified that descriptor to address expert concerns. The TeamUP rubric (Version 4) can be considered to be well validated at that level of consensus. The final rubric reflects underpinning theory, with no areas of conceptual overlap between rubric domains.

Conclusion: The final TeamUP rubric arising from this study validly measures individual student teamwork skills and can be used with confidence in the university setting.

KEYWORDS
Teamwork; Educational assessment; Rubric; Undergraduate; Validation study; Expert opinion; Modified Delphi technique
**BACKGROUND**

The Teamwork assessment rubric evaluated in this paper is designed to test an essential ‘soft’ employability competence desired by the vast majority of surveyed employers (Hart Research Associates, 2009; Robles, 2012). The Australian Qualifications Framework (2013) requires that higher education institutions develop students’ ‘Generic Skills’ including ‘working with others’ and ‘interpersonal skills’. Despite this readily articulated need, teamwork skills are usually not explicitly taught and rarely tested (McNair, 2005; Oliver, 2011). Within healthcare, poor teamwork skills are cited as causal factors in adverse outcomes (Centre for Maternal and Child Enquiries, 2011; Department of Health and Ageing, 2009; Douglas, Fahy, & Robinson, 2001; Garling, 2008; Manser, 2009).

Although nurses and midwives provide much of their care on a one-to-one basis they are also most often constituted in standing, discipline-specific teams - such as the nursing team on a shift at the ward level or the birth suite midwifery team. Additionally, nurses and midwives are part of multidisciplinary teams such as ‘the Renal Team’ or the ‘Maternity Care Team’ that include doctors and allied health professionals. Thus teaching and assessing the teamwork skills of health professionals deserves careful attention from health professional educators.

Some may argue that communication and teamwork abilities are inherent individual characteristics and therefore potentially unchanging (Belbin, 2010; Kinlaw, 1991). In contrast, this paper proposes a skills-based approach to teamwork skills development. This approach is consistent with a functional theoretical model that is the dominant contemporary theory (Burke et al., 2006; Hughes & Jones, 2011; Morgeson, DeRue, & Karam, 2010; Mumford, Van Iddekinge, Morgeson, & Campion, 2008). In 2012 we introduced academic teamwork assignments and the peer marking of individual teamwork skills in the Bachelor of Midwifery. An evaluation survey of these students indicated that students wanted specific guidance about teamwork skills (Parratt, Fahy, & Hastie, 2014). A literature review on teamwork skills and how they might be best assessed found no relevant, detailed, assessment rubrics were published (Hastie, Fahy, & Parratt, 2014). Following recommendations suggested by Hughes and Jones (2011) we modified the generic American Association of Colleges and Universities’ (ACCU) rubric (Rhodes, 2010) to create the TeamUP assessment rubric. The TeamUP rubric has undergone evaluation and further development in the subsequent years. The educational theory underpinning the development of the TeamUP rubric assessment process is described elsewhere (Hastie et al., 2014).

In 2013 we formed an Action research group of academics to systematically assess, plan, implement and evaluate our efforts to teach and assess teamwork skills. Currently there are nine core members in the Action research group; one in the discipline of education, one in management and seven health professional educators; three are midwives, two are nurses and two are engaged in complementary medicine. We designed the TeamUP educational enhancement and implemented it in all three years of the Bachelor of Midwifery and in one semester of the fourth year of the Bachelor of Education. Table 1 provides the key theoretical definitions; other key TeamUP terms have been modified during this research and are defined later in the paper.
### Table 1: Definitions of key terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behaviour</td>
<td>Behaviour is the internally coordinated responses (both actions and inactions) of the whole organism to internal and/or external stimuli, excluding responses more easily understood as developmental change (Levitis, Lidicker Jr, &amp; Freund, 2009).</td>
</tr>
<tr>
<td>Learning</td>
<td>Learning is the process of changing behaviour as a result of experience. Consistent with constructivist learning theory, when students are facilitated to engage in teamwork projects, individual learning occurs in communities of practice with the aim of enhancing socialisation for all team members (Smith, 2003, 2009).</td>
</tr>
<tr>
<td>Rubrics</td>
<td>A rubric is a scoring tool that divides an assignment into its component parts and objectives, and provides detailed descriptions of what constitutes acceptable and unacceptable levels of performance (Hastie et al., 2014; Stevens &amp; Levi, 2004).</td>
</tr>
<tr>
<td>Skill</td>
<td>A skill is the ability to perform a specific behaviour well. A skill requires knowledge, attitude and practice; skills develop over time (Yallop et al., 2005).</td>
</tr>
</tbody>
</table>

#### Teams:

1. **Standing teams**
   - Standing teams (e.g. disciplinary or multidisciplinary teams) are relatively stable and persist over time so there is an expectation of ongoing relationships and interactions between members often exceed role functions (Alge, Wiethoff & Klein 2003).

2. **Drill teams**
   - Drill teams (e.g. emergency or surgical teams) continue for only short periods and membership changes; however roles are clearly defined and as a result member interactions are usually limited to role functioning (Alge, Wiethoff & Klein 2003).

TeamUP is a package of teaching and learning activities based on a theoretically grounded assessment rubric (Hastie et al., 2014). This rubric was designed to guide student learning and to assess the teamwork skills that they practice while undertaking team-based academic assignments in higher education. The focus of the TeamUP rubric is on the fundamental teamwork behaviours that can be taught, practised and assessed so that individual students are enabled to develop their skills over time. Students use the rubric to provide anonymous peer feedback to each other; the subject coordinator then assigns individual teamwork marks, taking into account peer feedback and other evidence such as project plans and meeting minutes. The other elements of TeamUP are six lectures and six associated skills practice tutorials on topics directly relevant to the skill domains referred to in the rubric.
Validity of teamwork performance assessment needs to be addressed systematically because of the potential consequences for graduates, and ultimately patients, of unsound practices. This paper reports on research aimed at strengthening the validity of the TeamUP rubric. The question guiding this validation study was: to what extent do experts agree that the TeamUP rubric is a valid assessment of ‘teamwork skills’?

According to traditional psychometric theory, “validity pertains to the extent to which a test measures what it is supposed to measure” (Schuwirth & van der Vleuten, 2011, p. 786). Validity has traditionally been considered as criterion, content, and construct related. Messick (1995), a well-respected expert in assessment validity, argues that ‘validity’ is not a property of a test (i.e. the test itself is not valid). He instead maintains that the concept of validity reflects the meanings that are made of assessment results and what is done based on those meanings (e.g. passing versus failing a student). Likewise, Schuwirth & van der Vleuten (2011) claim that the best way to consider validity is “an overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations” of assessment outcomes (p. 741).

This paper focuses on content validity by evaluating the representativeness and relevance of the items in a tool or assessment (Lynn, 1986; Polit & Beck, 2006). A tool with good content-validity is necessary for supporting an argument of construct validity. Construct validity is defined as “the extent to which empirical evidence confirms that an inferred construct exists” (e.g. ‘teamwork skills’) “and that a given assessment procedure is measuring the inferred construct accurately” (Popham, 2011, p. 89).

**METHODOLOGY**

The research reported here relates to the evaluation phase of an ongoing action research project called TeamUP. Action research is a widely used methodology where co-researchers repeatedly undertake cycles of assessing, planning, implementing and evaluating, which in turn generates practical knowledge that can change practice (Reason & Bradbury, 2008). The practice change we are seeking is our own academic practice as teachers and assessors of student teamwork skills. The overall TeamUP project is grounded in feminist methodological commitments. The Delphi study reported here gathered both quantitative and qualitative data.

The Feminist nature of the action research depends upon the definition that; feminism is the theory, research and practice that aims to identify, understand and change intrapersonal and social factors that sustain women’s disempowerment (Harrison & Fahy, 2005). In the case of nursing and midwifery students we wanted to promote their empowerment by developing teamwork skills that will enable them to be both wiser and more empowered in their teamwork interactions; particularly in the workplace. In line with our feminist commitment, previous and ongoing aspects of the TeamUP project have specifically included engagement with the students themselves about their experiences of teamwork assessment using the TeamUP rubric (Hastie et al., 2014; Parratt et al., 2014); a paper reporting a qualitative evaluation of student experience is currently under review.

As required in a validation study, we aimed to explicitly clarify and synthesise experts’ opinions – including, but not imposing, our own – in a way that is methodologically defensible. Feminist
researchers have a commitment to explicitly and transparently consider their own relevant knowledge and experiences in the research process (DeVault & Gross, 2007;Etherington, 2004; Harrison & Fahy, 2005; Kirsch, 1999). This study honoured that commitment by using Feminist Collaborative Conversations; these conversations have been advocated to improve the quality and transparency of qualitative data interpretation (Capobianco, 2007; Dunn, 2010; Feldman, 1999; Hollingsworth, 1992; Hoskyns, 2011; MacPherson, 2010). Conversations included critical discussion about the theoretical and philosophical basis of our own and participants’ ideas (Feldman, 1999; Hastie et al., 2014; Hollingsworth, 1992).

**Research Design**

We used a Delphi design because it is an effective method for gaining consensus where available literature is insufficient, overwhelming or contradictory (Fletcher & Marchildon, 2014; Jones & Hunter, 1995). Delphi’s strength is the ability to solicit expert opinions and subsequently synthesise these into new knowledge based largely on consensus. Delphi method has diverse applications which are widely used (Vernon, 2009), for example in health education (Fraser, 1999), content standard validation (van der Schaaf & Stokking, 2011), teamwork studies (Clay-Williams & Braithwaite, 2009), and rubric development (Newman, Lown, Jones, Johansson, & Schwartzstein, 2009). While survey methods are often used, Delphi is a flexible research approach with many possible modifications (Jones & Hunter, 1995; Vernon, 2009). Researchers report Delphi modifications such as the number and timing of Delphi rounds: a series of rounds may be held over weeks or months or they may occur over a single day (ie a ‘Delphi conference’) (Vernon, 2009). There are also computer-based approaches that do without rounds altogether and enable varying modes of participation (Gordon & Pease, 2006; Vernon, 2009). One modification we made was to allow participants to respond either by online survey or via interviewer-guided survey by phone, Skype or face-to-face. In awareness of the time commitment of busy professionals who participated in this study, we also modified the usual Delphi approach (retaining participants through all rounds) so that the rounds continued as new participants were recruited as needed. Delphi panel size and composition is known to vary according to individual study nuances (Keeney, Hasson, & McKenna, 2011). Homogeneous panels can have as few as five participants, but such panels are also likely to require (and attain) 100% consensus (Lynn, 1986). Heterogeneous panels, such as the one for this study, benefit from the diversity of member experience and opinion, thus relatively high numbers of participants are recommended to capture that diversity (Keeney et al., 2011). Attaining consensus, however, becomes more difficult as participant numbers increase (Keeney et al., 2011).

**Ethics**

The University’s Ethics Committee granted consent before the study commenced. Consent was assumed when a participant responded affirmatively to the invitation email and agreed to participate. Participation was voluntary and responses were de-identified. In order to recognise the contribution that the participants made to the TeamUP rubric and to demonstrate their accountability, the participants who agreed to be listed are named in the Acknowledgements.

**Recruitment**

We aimed to recruit professionals with recognised expertise in teamwork, leadership or communications. We define ‘professional’ as a person who exhibits high standards of ethics, is engaged in an activity for gain/compensation and has achieved mastery of their discipline (“Professional 1,” 2012). Using phone and email, we recruited a heterogeneous group of
participants through the professional networks of the TeamUP Action research project. Further professionals were recruited by referral from participants. Evidence of their expertise included: relevant books/paper publications, complex team management experience, paid/unpaid work involving teamwork, employment as leadership and/or teamwork consultant. Invitations, including information sheets, were emailed to sixty-six professionals of whom thirty-five agreed to participate. Non-responders were emailed a single reminder.

In deciding the number of participants for the expert panel we strove for a balance so that both heterogeneity and a reasonably high level of consensus could be achieved. The panel comprised 35 professional members, predominantly Australian, with some from the United Kingdom and United States. The professions represented included nursing, midwifery, speech pathology, medicine, education, engineering, management, international and community development and tourism. Our expert participants were anonymous to each other and the opinions of individual participants were not available to other participants. This minimized the risk of an individual’s opinions influencing the opinions of others (Jones & Hunter, 1995).

The Survey

The psychometric theory underpinning validity testing, discussed above, guided the development of the survey used in the Delphi rounds (Messick, 1995; Popham, 2011; van der Schaaf & Stokking, 2011). A unique three-part survey was used in each of the Delphi rounds. Part 1 of all surveys was based on the current version (at that time) of the TeamUP rubric (i.e. Versions 1-3). Items on the rubric consisted of the ‘teamwork skills’ definition, the definitions of the five domains, and the behavioural descriptors within each domain.

In responding to the survey participants were asked:

a) whether they agreed with the definitions of teamwork skills and its five domains;

b) to give their opinion about these definitions;

c) to validate each descriptors under each domain by answering ‘YES’ or ‘NO’ to the question:

"Does this criterion appropriately measure a skill necessary for effective teamwork that is so important that, if unmastered, is likely to result in ineffective or dysfunctional team interactions in the work setting?"

d) to rephrase, delete, add or suggest alternative descriptors if they thought that necessary.

Part 2 of the survey asked participants for further comments about the rubric overall, including recommendations for improvement.

Part 3 asked three questions: consent to include name and affiliations in an acknowledgments list; willingness to be emailed an invitation for follow-up, and; recommendations of other professionals skilled in teamwork. In the online survey, a fourth question asked about participants’ willingness to be contacted by a researcher to discuss responses.

Data collection

Round 1 of data collection began with TeamUP rubric Version 1 (Hastie et al., 2014). We conducted three Delphi rounds between May 2013 and April 2014, revising the rubric after each round. This number of Delphi rounds is consistent with Keeney, Hasson and McKenna (2011) recommendations
for heterogeneous panels such as ours. Most participants critiqued the rubric by online survey. Participants contributing via interviewer-guided surveys were emailed a copy of the survey prior to interview. All members of the research team conducted one or more of the fourteen interviewer-guided surveys. Interviewer-guided surveys were undertaken to enable participants to express and clarify their opinions with ease. For the same reason we added the option of a brief telephone discussion for the remaining participants undertaking the survey online; two participants took up this option.

Round 1 ended after only six participants had given feedback because data analysis (which we conducted concurrently with data collection) indicated that further rubric refinement would be necessary. Twenty-six new participants were recruited in Round 2, and a total of thirteen participants completed Round 3. Round 3 comprised four participants who had completed Round 1, six from Round 2 and three who were newly recruited. We stopped collecting data after Round 3 because responses had dwindled and data saturation was apparent (i.e. no new opinion was emerging). Thus, the three Delphi rounds produced rubrics Versions 2 and 3 and ultimately, the final TeamUP rubric (Version 4).

Data management
Following each Delphi round the first author collated the quantitative and the qualitative data (i.e. expert comments) into a table organised according to the rubric items. She circulated this table to the research team for their consideration. The aim was to revise the rubric and strengthen validity of subsequent rubric versions. Research team members responded both verbally and in writing. In the sections below the process for quantitative analysis is described first and then the qualitative analysis is presented.

Quantitative Analysis

In designing the data analysis we primarily followed Polit & Beck (2006). These methods are consistent with the recommendations of Popham (2011) and the Standards for Educational and Psychological Testing (American Educational Research Association (AERA), American Psychological Association, & National Council of Measurement in Education, 1999). In this section we refer to various iterations of the TeamUP rubric (Versions 1-3) as ‘the scale’. We engaged in analysis after each Delphi Round. Quantitative analysis involved three calculations:

1) the Item Content Validity Index (I-CVI) for each definition and behavioural descriptor (Polit & Beck, 2006);
2) the Domain Content Validity of each of the five teamwork domains (D-CVI/Ave), which is an average of the item content validity indices for the domain definition and each descriptor within each domain, and;
3) the overall Scale Content Validity Index (S-CVI/Ave), which is an average of the item content validity indices of the whole scale including all the definitions (Polit & Beck, 2006).

To determine the I-CVI we added all ‘YES’ answers for each individual item then divided it by the total number of experts who responded to that item - the result is expressed as a percentage. A single item may be deemed valid when its I-CVI reaches a pre-set level of consensus (i.e. agreement between experts) (Jones & Hunter, 1995; Lynn, 1986). Pre-set consensus levels in other studies vary between 51% and 80%; usually the larger the number of participants, the lower the level of
consensus expected (Green, Jones, Hughes, & Williams, 1999; Keeney, Hasson, & McKenna, 2006). Recognising that the larger and more diverse the panel the harder it is to achieve consensus, we preset item consensus level at 70%; this level is considered ‘strong’ for this type of panel (McIlrath, Keeney, McKenna, & McLaughlin, 2010, p. 272).

The D-CVI/Ave was calculated for each domain by summing the I-CVI (expressed as a percentage) and then dividing that figure by the number of items in that domain. We made the decision to calculate this sub-scale D-CVI/Ave measure in order to gain a richer picture of the overall rubric. To calculate the S-CVI/Ave we summed all the I-CVIs and divided that number by the total number of items in the scale. This averaging method for determining S-CVI/Ave was chosen because it more readily addresses the possibility of a chance inflation of expert agreement or disagreement than do other methods (Polit & Beck, 2006). The S-CVI/Ave is a more generous measure so we aimed for 10% above our pre-set item consensus level, ie S-CVI/Ave = 80%; this level is lower that Polit and Beck’s (2006) recommended level to allow for our larger heterogeneous panel.

**Interpretation of Qualitative Data**

In all validation studies the researchers make decisions about how to synthesise the content of items on each iteration of the scale being evaluated by experts (Messick, 1995; Polit & Beck, 2006; van der Schaaf & Stokking, 2011; Vernon, 2009). The psychometric methodology literature lacks any guidance of how to synthesise the content of items on each iteration of the scale. The types of teams varies across contexts; e.g. business, various contexts of health care, community service etc; each type of teamwork can only be well understood in its specific context (Valentine, Nembhard, & Edmondson, 2015). As outlined above, we had a very diverse group of experts, most of whom were not teaching undergraduate teamwork skills. Three members of the research team conducted the majority of the interpretation of qualitative data. These academic researchers were expert in teaching and assessing teamwork and were very familiar with the TeamUP rubric and its usefulness as both a teaching and assessment tool in the context of undergraduate education. Interpretation of data proceeded via Feminist Collaborative Conversations. In these conversations we sought to consider:

i) the experts quantitative agreement with each item and any qualitative comments in light of their score;

ii) how individual expert opinion related to the words written by other experts;

iii) what modifications and additions needed to be made to the rubric;

iv) how the items would be understood by the students who are the recipients of the assessment rubric.

During Feminist Collaborative Conversations we privileged expert opinion by directly referencing, discussing, comparing and contrasting their qualitative responses. The conversations continued until there was no further debate among the researchers (Reinharz & Kulick, 2007). At this point, the research team agreed that there was sufficient consensus among the panellists about the definitions and descriptors on the rubric. Equally, we agreed that the definitions and descriptors would serve their pedagogical purposes. This process continued over many months and enlarged to include the whole research team. While revising Version 3, we returned to the data from all three Delphi rounds and re-considered all the expert opinions. This was done to strengthen the trustworthiness of the
qualitative analysis and interpretation. All key participant concerns were appropriately incorporated in the final version of the rubric. We maintained a data trail of decision-making and rationales.

RESULTS

Figure 1 shows the levels of consensus among experts for each of the items in various versions of the rubric; for example, 19 items achieved 100% consensus.

[INSERT FIGURE 1 HERE, see below]

The S-CVI/Ave for the rubric was 92% for Version 2, and 89% for Version 3. The D-CVI/Ave for the five domains of Version 3 are shown in Table 2. This sub-scale analysis provided a more specific measure than the S-CVI/Ave scores indicating less agreement among panel members in Domain 4 (D-CVI/Ave = 77%). In this domain (Managing Conflict) some experts, for example, believed that only team leaders should learn about conflict management; yet in the pedagogical context of TeamUP all students are expected to learn conflict management. Participants often responded to the survey by qualifying their ‘YES’ answers with a comment. In our interpretation, participants were saying that the wording in the rubric was on the right track but further refinement was necessary. These refinements would have been missed if we had focused only on measuring the content-validity indices. Consistent with other Delphi researchers, we acknowledge the value in thinking about and integrating as much of expert opinion as we could, rather than striving for 100% consensus on each item (Fraser, 1999; Vernon, 2009). We therefore interpreted each I-CVI in tandem with the qualitative comments.

<table>
<thead>
<tr>
<th>Term</th>
<th>Version 3 Definitions</th>
<th>Changes*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teamwork Skills</td>
<td>A group of learned behaviours that provide the individual with the capability of working towards team goals whilst maintaining their own integrity.</td>
<td><em>A set of behaviours that can be learnt which provide the individual with the capability of working towards team goals whilst maintaining their own integrity.</em></td>
</tr>
<tr>
<td></td>
<td>I-CVI: 92%</td>
<td></td>
</tr>
<tr>
<td>Domain 1: Project Planning Skills</td>
<td>The planning skills that enable team members to work in synergy to produce a project plan of the best possible quality.</td>
<td>The actions indicating the team member has worked in synergy with others to plan a high quality project</td>
</tr>
<tr>
<td></td>
<td>D-CVI/Ave: 93%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I-CVI: 92%</td>
<td></td>
</tr>
<tr>
<td>Domain 2: Fostering a Team Climate</td>
<td>The emotional and social skills to foster a sense of trust and inclusiveness for each team member.</td>
<td>The actions fostering a sense of trust and inclusiveness for each team member</td>
</tr>
<tr>
<td></td>
<td>D-CVI/Ave: 91%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I-CVI: 91%</td>
<td></td>
</tr>
<tr>
<td>Domain 3: Facilitating the Contributions of Others</td>
<td>The skills to ensure that the processes of team interaction are effective in progressing the project plan; this means that accurate, relevant information is exchanged, understood and used by all team members.</td>
<td>The <em>actions ensuring</em> the processes of team interactions are effective in progressing the project plan.</td>
</tr>
<tr>
<td>D-CVI/Ave: 89%</td>
<td>I-CVI: 89%</td>
<td></td>
</tr>
</tbody>
</table>

| Domain 4: Managing Conflict | The personal and interpersonal skills to prevent, recognise, and address conflict in ways that strengthen overall team cohesiveness & effectiveness. | The *actions preventing, recognising, and/or addressing* conflict in ways that strengthen overall team cohesiveness & effectiveness |
| D-CVI/Ave: 77% | I-CVI: 77% |

| Domain 5: Contributing to the Team Project | The skills required to make a high-quality, individual contribution to the team project. | The *actions demonstrating that the team member has made* a high-quality, individual contribution to the team project |
| D-CVI/Ave: 93% | I-CVI: 93% |

**KEY:**
*Changes are indicated in italics and underlined*
I-CVI – the item content validity index for the domain definition
D-CVI/Ave - an average of the item content validity indices for the domain definition and each descriptor within each domain
The refinements that occurred when taking qualitative comments into account were used to revise Version 3 of the rubric. Table 2 ‘Definitions of Teamwork Skills and Domains’ shows how definitions that were essentially agreed upon by panel members were improved following qualitative data interpretation. Similarly, Table 3 ‘Teamwork skills Domain descriptors’ shows Version 3 domain descriptors along with the changes and additions made in response to the qualitative data. These tables also show the I-CVI for the six Version 3 definitions (Table 2) and the Version 3 domain descriptors (Table 3).

### Table 3: Teamwork skills domain descriptors

<table>
<thead>
<tr>
<th>I-CVI (%)</th>
<th>Domain name &amp; Version 3 Descriptors</th>
<th>Changes* and additions (n=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Project Planning Skills</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>a) Electing and supporting a project manager</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>b) Defining and agreeing on team goals and objectives</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>c) Defining and agreeing on quality standards for each part of the plan</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>d) Contributing to the development of the plan</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>e) Setting and agreeing realistic timeframes for each part of the plan</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>f) Participating in role allocations based on individual skills and learning needs</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>g) Willingly taking on a team role that can be completed on time to a quality standard</td>
<td></td>
</tr>
<tr>
<td>g) Agreeing to undertake a team role/s, to meet agreed quality standards and to complete on time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h) Participating in changing the plan if contingencies arise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Fostering a Team Climate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>a) Exhibiting an open, gentle, polite and friendly manner</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>b) Demonstrating self-awareness and emotional regulation</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>c) Demonstrating sensitive awareness of the feelings of others (including interpreting body language)</td>
<td></td>
</tr>
<tr>
<td>a) Exhibiting an open, polite and friendly manner</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Changes indicated by participants and panel members*
100  d) Actively contributing to team discussions
100  e) Listening attentively without interrupting
100  f) Cooperating with others to achieve project goals
92   g) Following up with others when there is concern about their feelings or contribution
92   h) Showing respect for the contributions of others (even if in disagreement)
83   i) Expressing genuine gratitude and praise generously

3. Facilitating the Contributions of Others

77   b) Taking turns at leading/coordinating a team meeting (including agenda preparation)
77   c) Taking turns at keeping and distributing brief meeting minutes (with action items and deadlines)
100  d) Leading and/or participating in teambuilding processes
100  e) Establishing and honouring team ground-rules
100  f) Ensuring that decisions are made in a timely manner
92   g) Listening attentively to the person who is speaking
70   h) Using an agreed signal and waiting to be called upon before speaking
92   i) Participating in consensus-building decision-making
92   j) Inviting other team members to contribute

k) Assisting team members if they are struggling to meet their requirements

4. Managing Conflict

a) Balancing the desire to maintain social harmony with the need to address potential or actual conflict
b) Being appropriately assertive: neither dominating, submissive, nor passive aggressive

c) Minimising unnecessary conflict by project planning and management

d) Approaching conflict with the aim to de-escalate

e) Participating in activities aimed at turning conflict into learning

f) Staying focused on the overall team goal/s

g) Challenging team processes that are not conducive to the achievement of team goals

h) Giving timely and constructive feedback on the behaviour of others

i) Being open to receiving and reflecting upon constructive feedback of own behaviour

5. Contributing to Team Project

a) Submitting assigned work at the agreed quality standard

b) Submitting assigned work within the agreed timeframe

c) Appropriately critiquing the work of others

d) Working to integrate the output of team members into the project

e) Demonstrating sufficient technological skills

f) Demonstrating relevant content knowledge

g) Adhering to appropriate academic standards (ie. writing, intellectual property, confidentiality, research ethics, social networking)

h) Evaluating the quality of the whole project and making needed changes

KEY:
*Changes are indicated in italics and underlined
I-CVI – the item content validity index for each Version 3 descriptor

The final TeamUP rubric (Version 4) was constructed from Version 3. Tables 2 and 3 and show the changes and additions which occurred following each iteration. Figure 2 ‘Teamwork skills in final TeamUP rubric’ shows the content of the final TeamUP rubric organised under the domains.
**DISCUSSION**

The study used both quantitative and qualitative methods to evaluate the validity of the TeamUP rubric. This Delphi study began with Version 1 of the TeamUP rubric which is grounded in the theoretical literature relevant to each domain of teamwork skills (Hastie et al., 2014). Having a tool that is well grounded in theory provides some evidence of the construct validity of the rubric (Messick, 1995) and thus Version 1 already had a degree of validity. This Delphi study answers the question: “To what extent do experts agree that the TeamUP rubric is a valid assessment of teamwork skills?”. We followed methodological guidance from both psychometric theory and Delphi research techniques. We added to this a methodologically rigorous and transparent synthesis of expert opinions to create a final version of the TeamUP rubric as a valid expression of ‘teamwork skills’ within an undergraduate setting.

During consensus development, participant comments often indicated an agreement in principle coupled with the need to further refine or clarify the language used; this was why the S-CVI/Ave dropped 3% between Version 2 and 3. After Delphi Round 2, we addressed participant identified areas of conceptual overlap between domains; we re-wrote definitions and moved items. In this refinement process we deleted and added some descriptors with an overall increase in items from 43 to 47 in Version 3. Similarly, our refinement of Version 3 produced the TeamUP rubric (Version 4) of 51 items. We added a single descriptor to Domain 1 and two to Domain 3 to capture participant meanings; the Domain 3 additions also reflected the theoretical underpinnings of this domain (Dick, 1991). The fourth descriptor, which emerged from our collaborative conversations, was added to Domain 4. This additional descriptor reflects that students often avoid conflict in an effort to maintain relationships, yet that comes at a cost of misalignment with their own values (e.g. fairness, truth or quality).

The overall S-CVI/Ave of 89% for Version 3 of the TeamUP rubric is well above our aimed-for target of 80%. The final TeamUP rubric has the strengths of Version 3 but also addresses the expert opinions given during Delphi Round 3. The rating system we followed, recommended by Popham (2001), was based on a categorical yes/no assessment by experts, rather than a Likert scale which allows more nuances of opinion. We did, however, note that when Likert scales have been used (such as by Lynn, 1986), they are likely to be four-point scales, which are subsequently dichotomised. Using a dichotomous rating scale may possibly inflate the level of inter-rater consensus and create discrepancy between the scores and the expert comments. We believe we have responded to this perceived weakness in method because we carefully considered all expert comments in the creation of the final version of the rubric. Another possible limitation is that some experts only too part in one Delphi round, we recruited experts through all three rounds; however, the diversity of opinion is also a strength. The Delphi method has been criticised for neglecting diverse opinion and encouraging conformity, which in turn purport to force consensus (Fletcher & Marchildon, 2014). Overall, we believe our modified Delphi method addresses these criticisms.

Delphi researchers have an influential role during consensus development where qualitative comments form part of the research data. This influential role has been cited as a potential form of bias that is most likely to occur when researchers are opaque about how they create their final version (Green et al., 1999; Vernon, 2009). Our engagement in Feminist Collaborative Conversations ensured transparency because these conversations required us to explain, listen, critique and synthesise knowledge that privileged the experts’ intended meanings (Harrison & Fahy, 2005;
Reinharz & Kulick, 2007). We consider that the wording changes we have made to individual items has strengthened them and that these changes were consistent with the qualitative comments of the experts.

Ultimately, and consistent with Messick (1995), in future we intend to evaluate the construct-validity of the Final TeamUP rubric by seeking an overall evaluative judgement from employers of our new graduates compared with students graduating from most other program that currently do not systematically teach and assess teamwork skills. Further research is currently being completed to evaluate the TeamUP experience of midwifery students and new graduates who have been exposed to one or more years of the TeamUP educational enhancement.

**CONCLUSION**

In this paper, validity is considered, according to Messick (1995), as the “overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations” of assessment outcomes (p. 741). Our research strengthens the evidence that the TeamUP construct of Teamwork Skills is valid for the purpose of teaching and assessing individual student teamwork skills in an academic environment. This paper provides the empirical evidence from experts while the theoretical rationales have been published (Hastie et al., 2014) and somewhat modified in this Delphi study, as described above. Our claim to validity is consistent with Messick (1995) and Schuwirth & van der Vleuten (2011) who assert that the best way to consider validity is to see it as a subjective, qualitative judgment, normally taken by someone with expertise in the domain being tested. Academics can now use the TeamUP rubric to teach and assess teamwork skills with confidence that they are using a well validated tool to do so.
REFERENCES


ACKNOWLEDGEMENTS

The following expert participants in this modified Delphi project agreed to be acknowledged:

- Professor Jill Thistlethwaite, University of Queensland;
- Professor Judith Lumby, University of Adelaide;
- Dr Sharon Mickan, University of Oxford, UK;
- Professor Kristine Martin McDonald, Victoria University;
- Dr Lyn Gardon, NSW Department of Education and Communities;
- Dr Robyn Clay-Williams, University of New South Wales;
- Dr David Solnet, University of Queensland;
- Dr Linda Crane Assoc. Dean Quality Learning & Teaching, Bond University;
- Michelle Bradley RN GDipSE MSc Clinical Nurse Manager Mental Health, South Eastern Sydney Local Health District;
- Anne Saxton Conjoint Associate Professor, University of Newcastle;
- Rae Conway RN MHN BHSc MBA Manager Nursing and Midwifery Workforce Development and Leadership, Northern NSW Local Health District;
- Nicola Graham, Speech Pathologist;
- Professor Michelle McLean, Faculty of Health Sciences and Medicine, Bond University;
- Professor Anneke Fitzgerald, Health Management Griffith Business School;
- Annette Symes Executive Director, Nursing and Midwifery, Northern NSW Local Health District;
- Dr Phil Connors Course Director Humanitarian Assistance, Deakin University;
- Dr Catherine McLoughlin Australian Catholic University;
- Professor Soo Downe, University of Central Lancashire, UK;
- Fiona Faulks Manager Women’s & Children’s Services Bendigo Health;
- Dr Gary Hamlin PhD Faculty of Health Sciences and Medicine, Bond University;
- Dr Kathy Eljiz, Griffith University; Dr Louise Kippist School of Business, University of Western Sydney;
- Dr Catherine McDermott Faculty of Health Sciences and Medicine, Bond University;
- Dr Peggy L Chinn Professor Emerita of Nursing, University of Connecticut USA;
- Matthew W. Ohland, Professor of Engineering Education, Purdue University USA and Professorial Research Fellow, Central Queensland University;
- Margaret Kelly RN, Manager Acute and Aged care.
Figure 1: Level of consensus for items in TeamUP rubric Version 1, 2 and 3.

- Version 3, n=47 items, 13 participants (Round 3)
- Version 2, n=43 items, 26 participants (Round 2)
- Version 1, n=43 items, 6 participants (Round 1)
Expert Validation of a Teamwork Assessment Rubric: A Modified Delphi Study

Figure 2: Teamwork skills in final TeamUP rubric

Teamwork skills: A set of behaviours, that can be learnt, which provide the individual with the capability of working towards team goals whilst maintaining their own integrity.

Project Planning
The actions indicating the team member has worked in synergy with others to plan a high quality project.

- Eliciting and supporting a project leader
- Defining and agreeing on team goals and objectives
- Defining and agreeing on quality standards for each part of the plan
- Contributing to the development of the plan
- Setting and agreeing realistic timeframes for each part of the plan
- Participating in role allocations based on individual skills and learning needs
- Agreeing to undertake a team role/s, to meet agreed quality standards and to complete on time
- Participating in changing the plan if contingencies arise

Facilitating the Contributions of Others
The actions ensuring the processes of team interactions are effective in progressing the project plan.

- Exchanging accurate, relevant information that can be understood and used by all team members
- Taking turns at leading/coordinating a team meeting (including agenda preparation)
- Taking turns at keeping and discontinuing brief meeting matters (with action items and deadlines)
- Leading and/or participating in team building processes
- Establishing and honouring team ground rules
- Ensuring that decisions are made in a timely manner
- Listening attentively to the person who is speaking
- Using an agreed signal and waiting to be called upon before speaking
- Participating in consensus building decision making
- Involving other team members to contribute
- Assisting team members if they are struggling to meet their requirements

Fostering a Team Climate
The actions fostering a sense of trust and inclusiveness for each team member.

- Exhibiting an open, polite and friendly manner
- Demonstrating self awareness and emotional regulation
- Demonstrating sensitive awareness of the feelings of others (including interpreting body language)
- Actively contributing to team discussions
- Listening attentively without interrupting
- Cooperating with others to achieve project goals
- Following up with others when there is concern about their feelings or contributions
- Showing respect for the contributions of others (even if challenging or expressing disagreement)
- Expressing genuine praise and gratitude as a regular part of behaviour

Managing Conflict
The actions preventing, recognising, and/or addressing conflict in ways that strengthen overall team cohesion & effectiveness.

- Minimising unnecessary conflict by project planning and management
- Being appropriately assertive: neither dominating, aggressive, nor passive: aggressive
- Balancing the desire to maintain social harmony with the need to address potential or actual conflict
- Approaching conflict with the aim to resolve it and maintain relationships
- Participating in activities aimed at turning conflict into learning
- Staying focused on the overall team goal/s
- Challenging team processes that are not constructive to the achievement of team goals
- Giving timely and constructive feedback on the behaviour of others
- Being open to receiving, reflecting and acting on constructive feedback of own behaviour

Contributing to the Team Project
The actions demonstrating that the team member has made a high quality, individual contribution to the team project

- Submitting assigned work at the agreed quality standard
- Submitting assigned work within the agreed timeframe
- Appropriately completing the work of others
- Working to integrate the output of team members into the project

- Demonstrating sufficient technological skills
- Demonstrating relevant content knowledge
- Adhering to appropriate academic integrity
- Evaluating the quality of the whole project and making needed changes