

Development of a Scoring System for the Team Effectiveness Questionnaire (TEQ)

Abstract

The purpose of this paper is to describe and document the development of a standard score (T-score) for the Team Effectiveness Questionnaire (TEQ). The TEQ was developed as an instrument to measure the effectiveness of engineering students working in teams. It is also useful in helping students understand their strengths and weaknesses as effective team players. The TEQ was developed around the premise that an effective teaming experience is obtained when team members perform well, behave as a team, and maintain a positive attitude toward teaming. The foundational premise of the TEQ is that an effective teaming process directly contributes to the overall effectiveness of the team. An effective teaming process is measured by the mastery of the following seven constructs: common purpose, goal clarification, role clarity, psychological safety, mature communication, productive conflict resolution, and accountable interdependence. In order to fully utilize the TEQ, a scoring system is necessary to facilitate calculating and interpreting results.

In choosing the norm reference standard score, T-score, we analyzed and evaluated several other scoring categories such as: content reference, criterion reference, and norm reference. The decision to use the T-score type was made after analyzing and evaluating the characteristics, objectives and/or intentions of the various types. The standard T-score was selected because: 1) it would not report negative values, 2) people are familiar with a 0-100 scale which would allow for easier interpretation, and 3) this type will allow individuals to compare their responses with the norm group.

The use of a standard score will minimize false interpretations that a non standardized score could lead to. Without a standard score the interpretation of the results will depend on the criteria of the person that has access to the results. The criteria and procedure used to develop the standard score will be described throughout this article, as well as the characteristics of the norm group that were used as a reference for the interpretation of future results. The mean and the standard deviation for each of the constructs are provided as well as an example of how to calculate a score and how to report it as a T- score. This standard score will be useful in accurately interpreting the individual's ability to perform as an effective team member and help them realize what their strengths and deficiencies are in order to improve those characteristics that an effective team player should have.

Introduction

The Effective Teaming Laboratory at the University of Nebraska-Lincoln developed the Team Effectiveness Questionnaire (TEQ) in 2001 to measure team effectiveness. It is composed of seven constructs considered to be necessary for effective performance of the team. The TEQ has three parts. The first part is used to collect demographic data and information on individual preferences regarding teams and previous experience in teams. The second, where scoring is the focus, is related to the seven constructs. Forty eight questions are asked in relation to the constructs. The third section asks questions about teaming issues and the level of familiarity each

individual has with them. The TEQ is an attempt to answer the question: Does the existence of specific team characteristics create a high performing team?⁷

In order to make this determination we must be able to obtain a score for the TEQ. Two main objectives were identified for developing the scoring system. The first objective is to be able to compare students' abilities with the seven constructs with a norm group. This comparison provides each individual the data to determine which constructs he/she is deficient in and will need to improve upon in order to be a part of an effective team.

The second objective is to measure students' ability in relation to the seven constructs at two different instances. The first instance is at the beginning of working on a team, which would capture their abilities on the seven constructs prior to the teaming experience and the second instance will be after the teaming experience. We anticipate that between the two instances students will undergo training which will improve their abilities to work on a team. These two measures will lead to a comparison of the two measures analyzing how people improve their effectiveness during this program which should end in a better result of their teamwork skills.

Based on these two objectives we examined three categories of scores, content reference, criterion reference, and norm reference. In the content reference the individual is compare to the maximum score possible of the test. In the criterion reference the person is compare to what they know and what they can do. In the norm reference the individual is compare to others in an specific group. Of the three types the one that is applicable to our questionnaire is the norm- reference that allows individual scores to be interpreted with reference to others (the norm- group). The norm reference score considers different types of scores; the most common are the Rank scores and Standardized scores.⁴

The rank score only considers the higher and lower score avoiding those in the middle; in rank scores the shape of the original distribution is lost unless this distribution is normal. This implies that if we do not have a Normal distribution the report scores won't be as the original, they are not equal units of measurement. Yet the considerations before, rank scores are easy to calculate and to understand due to the similarity to percentages.

The standard Scores compare inter-individual performance considering mean and standard deviation. It keeps the original distribution of the raw scores that will help to go back and forth from one score to the other. Another consideration for these scores is that they show where the individual falls in a distribution. It allows a comparison of individual results or comparison between groups. The different standard scores are: z-score, Army General Classification Test (AGCT) Scores, College Entrance Examination Board (CEEB) Scores, Deviation Intelligence Quotient (IQ), and T -score The Z-score is not only one type of score but is also the basis to others scores. The limitation of this score is that it reports half of the scores as negative values such as -1, -2, -3. AGCT scores are use mainly for military purposes. The CEEB- scores as the AGCT are used for specific areas. This type of score has a mean of 500 and a standard deviation of 100 and is used by the Educational testing services. T-scores are based on Z-scores, but it considers a mean of 50 and a standard deviation of 10. These parameters won't allow a negative score, what makes it good for interpretations. On the other

hand this score could be confused with the T-scale score, which is not the same. These two scores are equal when the original distribution is a normal distribution.^{1, 2, 4, 5, 6}

The T- score is a standard and norm- reference linear score where comparisons are expressed as number of standard deviations between any specified score and the mean of the group of reference⁴. This type of score was chosen for multiple reasons. The first consideration was that a non-normal distribution can be converted freely from raw- score values to a linear standard score without changing the shape of the original distribution for being a linear standard score. This type of score retains the shape of the raw score distribution changing only the metrics. Secondly, these scores will not report negative values. The use of a standard score such as a T- score will help to interpret and compare individuals, other groups, and universities or colleges.⁴

The norming sample

We established a norm group with the purpose of comparing future results recognizing that it should be revised when more data is collected. The sample was compiled from a group of 987 students from Texas A&M University who were enrolled in the course Foundations of Engineering I.

A profile of the characteristics (gender, major, classification, and ethnicity) of the sample is shown. 20 % of the students were females, and 80 % male; 50% of the students were Aerospace, Civil and Mechanical and the other 50% were enrolled in the other engineering disciplines; 83% freshmen and 17% comprised of sophomore, juniors, seniors and graduate students; and 78% Caucasian (non-Hispanic) , 8% Hispanic/ Latino/ Mexican American, 3% Asian American, 2% International Student, 2% Multi-Racial, 2% African American/ Black (non-Hispanic), and 1% Alaska Native/ Native American

Being an engineering student is the main characteristic that future students taking the survey will share with the norm group. Another shared characteristic is the age range of the people taking the survey.

Reliability of the norm

The norms for each of the seven constructs were established by calculating the mean and the standard deviations for each question in the TEQ and the overall mean for each of the seven constructs. Reliability and correlations were checked for internal reliability of the instrument. These two are important to consider before any scoring system is created. The Reliability will tell if the instrument is consistent, reliable, and if the set of questions are returning a stable response. On the other hand, the correlations between the variables are important because these will help to understand the relationship between the constructs.

For the reliability, the Cronbach's alpha based on standardized items was chosen. This index is used to measure internal consistency. The Cronbach's alpha is associated with the variation accounted for by the true score of the underlying construct³. In the literature Cronbach's alpha indexes above 0.7 are considered to be reliable. The results for this sample are

shown in Table 1. The Cronbach's alpha based on standardized items for the overall TEQ was 0.933. Since the score is close to one, we can conclude that the TEQ is consistent and reliable. For those constructs that the Cronbach's alpha is below 0.7 future review of the questions will be made.

Table. 1 Reliability

Construct	Cronbach's alpha based on standardized items	Items per construct
Overall	.933	
1. Productive Conflict Resolution	.513	6
2. Mature Communication	.699	5
3. Role Clarity	.718	4
4. Accountable Interdependence	.746	4
5. Goal Clarification	.634	4
6. Common Purpose	.745	5
7. Psychological Safety	.532	5

As was mentioned above correlation explains the relationship between variables. Two variables are considered to be related with a coefficient of 0.5. The correlations obtained between the constructs are shown in Table 2. The constructs that are highly correlated are common purpose with goal clarity and accountable interdependence with a correlation coefficient above 0.7. A change in common purpose will imply a change in the other two constructs. On the other hand, psychological safety and accountable interdependence have the lowest correlation (0.408), which means that a change in psychological safety won't necessarily imply a change in accountable interdependence.

Table 2 Correlation Table

	Conflict	Mat. Com	Goal	Purpose	Psy Safety	Role	Mut. Acct.
Conflict							
Mat. Com	.563						
Goal	.512	.683					
Purpose	.558	.643	.746				
Psy Safety	.469	.555	.443	.476			
Role	.475	.603	.656	.664	.446		
Mut. Acct.	.545	.592	.620	.716	.408	.596	

After setting the sample, a standard score was established for future interpretation of the responses to the TEQ.

Calculation of the score

To interpret average ability in each scale, the raw scores of each student is going to be reported as a standard T- score with the following parameters for each construct, shown in Table 3

Table 3 Mean and Standard Deviation of the constructs

CONSTRUCT	MEAN	SD
1. Productive Conflict Resolution	3.5301	0.46235
2. Mature Communication	4.0501	0.49497
3. Role Clarity	3.9283	0.59397
4. Accountable Interdependence	3.711	0.73457
5. Goal Clarification	4.0564	0.59648
6. Common Purpose	3.8071	0.62503
7. Psychological Safety	3.6623	0.53206

An example of the computation of the T-score for the construct of conflict resolution is provided below. The following is the procedure to obtain the score of each person for this construct.

1. Obtain the average of the raw score of the questions of the TEQ that correspond to the construct of conflict resolution.
2. Transform the average into a T- score using the MEAN and the SD of the norm group (M= 3.5301, SD= 0.46235) with the conversion formula as follows.

$$T - SCORE = \frac{raw.sore - norm.mean}{SD} * 10 + 50$$

$$T - SCORE = \frac{raw.sore - 3.5031}{0.46235} * 10 + 50$$

As described, the procedure to calculate each one of the T- scores for the seven constructs uses the corresponding means and standard deviation. After computing the T-score for each construct the overall score can be obtained by finding the average of the T-score of the 7 constructs. This overall score tells the individual their average performance as an effective team player. Hypothetical abilities as an effective team player of a student reported as a T- score are being described in table 4.

Table 4 Scoring Example

CONSTRUCT	Average	T-Score
1. Productive Conflict Resolution	4.78	78
2. Mature Communication	4.00	49
3. Role Clarity	4.00	51
4. Accountable Interdependence	3.75	51
5. Goal Clarification	3.50	41

6. Common Purpose	3	37
7. Psychological Safety	4.20	60
8. Overall score	3.73	52

For this particular student the results obtained in the survey in comparison to the norm group shows that the student has a scores within the norm in predictive mature communication, role clarity, accountable interdependence; the student has scored below the norm in goal clarification and common purpose, but his/ her psychological safety and conflict resolutions are above the norm. Based on this data this student an overall has an average score compared to the norm.

Conclusion

When developing a survey with the purpose of measuring and comparing results, a standard score should be calculated. The type of score chosen depends on what kind of reference is going to be used for comparison, how the raw data is collected, and which is the desired way to report the data.

Several types of scores could be chosen but the most important thing to keep in mind is to avoid confusion among the interpreters (students, teachers, trainers, or others). Reporting scores that receive a negative value could lead to misinterpretations of the real results and could create negative impacts on the individual. The person choosing the type of score should have a clear understanding the message they want to give to those with access to the results, and those that will interpret them. The development of the standard score was important for future interpretations of the instrument even though it is important to consider that it should be revised in the future with more and diverse data. Also, it is important to consider that any change in the instrument will affect the standard score as well as those obtained by the students and its interpretation.

By developing this standard score the Effective Teaming Laboratory at the University of Nebraska Lincoln will be able to interpret the results obtained by the students taking the current version of the TEQ, and address the needs of training in specific constructs that might strengthen the development on the specific teaming skills that will lead the students to become effective team member of effective teams.

Reference

1. Brown, Frederick G. *Guidelines for Test Use: A commentary on the Standards for educational and Psychological Tests*. National Council on Measurement in Education. 198
2. Downie, Norville M. *Types of Test Scores* Houghton Mifflin Company. 1968 pp. 25- 45.
3. Hatcher, L. (1994). *A step-by-step approach to using the SAS(R) system for factor analysis and structural equation modeling*. Cary, NC: SAS Institute.
4. Lyman, Howard B. *Test Scores & What They Mean*. Fifth edition. New Jersey: Prentice Hall, Inc.
5. Nunnaly, J. (1978). *Psychometric theory*. New York: McGraw-Hill.
6. Monroe Miller, David. *Interpreting Test Scores* John Wiley & Sons, Inc. 1972 pp 20- 42.

7. Simon, Laura Carmen. Study of the Performance of Student Teams in Engineering Education. Lincoln, Nebraska .
2001 pp. 11-12.