



## OPEN ACCESS

## EDITED BY

Yoed Nissan Kenett,  
Technion Israel Institute of Technology, Israel

## REVIEWED BY

Maor Zeev-Wolf,  
Ben-Gurion University of the Negev, Israel  
Marzia Saglietti,  
Sapienza University of Rome, Italy

## \*CORRESPONDENCE

Nira Mashal  
✉ mashaln@mail.biu.ac.il

RECEIVED 08 March 2023

ACCEPTED 07 December 2023

PUBLISHED 03 January 2024

## CITATION

Kofman Talmy A and Mashal N (2024)  
Quantitative and qualitative initial validation of  
the accountable talk questionnaire.  
*Front. Commun.* 8:1182439.  
doi: 10.3389/fcomm.2023.1182439

## COPYRIGHT

© 2024 Kofman Talmy and Mashal. This is an  
open-access article distributed under the terms  
of the [Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted which  
does not comply with these terms.

# Quantitative and qualitative initial validation of the accountable talk questionnaire

Ayelet Kofman Talmy and Nira Mashal\*

Faculty of Education, Bar-Ilan University Ramat-Gan, Ramat Gan, Israel

**Introduction:** Accountable Talk (AT) has been extensively studied as a tool for improving argumentation and respectful discourse. While several research tools exist for evaluating AT, as of yet there no self-report assessment tool to measure AT directly, on a large-scale basis, and without significant costs. The aim of the current study was to develop and provide initial validation for a self-report AT questionnaire (ATQ).

**Methods:** One hundred students aged 11–12 years participated in the study. Exploratory Factor Analysis, content validity, and test-retest reliability were assessed. 50 students were also randomly assigned to the qualitative part of the study. These students were recorded while discussing in small groups a topic provided by the experimenter (animal research). The qualitative data was coded and then correlated with the quantitative data obtained from the self-report questionnaire.

**Results:** The results indicated that a 12-items questionnaire can reliably assess three separate and independent qualities of AT: accountability to a learning community, accountability to rigorous thinking, and accountability to accurate knowledge. The reliability of the ATQ was high with  $\alpha = 0.80$ . The Test-retest reliability was assessed at two time points separated by a 3-week interval with Pearson correlation. Excellent correlations ( $r > 0.98$ ) between the ATQ scores were found. The correlation coefficients between the three components of the self-reported ATQ and the observed data obtained from the group discussions were significant positive medium-high.

**Discussion:** We have shown that a concise 12-item questionnaire can assess the three main components of Accountable Talk within the framework of respectful discourse. The questionnaire showed good reliability and structural validity, with weak correlations between sub-topics supporting the distinction between different aspects of Accountable Talk. We suggest that the ATQ can be used to evaluate the effectiveness of intervention programs aiming to improve students' acquisition of Accountable Talk skills.

## KEYWORDS

accountable talk, argumentation, respectful discourse, assessment tool, self-report questionnaire

## 1 Introduction

Argumentation is a verbal and social process that consists of the construction of arguments containing explanations and justifications (Kuhn and Udell, 2003). Argumentation is an important skill for learning and teaching and as a tool for improving problem-solving processes, acquiring concepts, building knowledge, and fostering academic achievements (Kuhn et al., 1997; Lazarou et al., 2017; Webb et al., 2019; Hasnunidah et al., 2020). The intrinsic value of argumentation lies in its seamless integration into educational practices, ultimately leading to elevated profound learning experiences and more meaningful interpersonal interactions (Koichu et al., 2022). Argumentation extends beyond

formal debate; it encompasses everyday interpersonal endeavors to actively participate in discussions with others aimed at co-constructing knowledge and meaning (Bova and Arcidiacono, 2023). Argumentation also serves as a platform for interpersonal socialization (Bova and Arcidiacono, 2023). This aims to collaboratively build knowledge and meaning, facilitating a deeper understanding of decision-making, and negotiating positions (Bova and Arcidiacono, 2017). Argumentation requires a strict adherence to respectful discourse rules, which establish a framework to provide participants with the knowledge and skills needed for engaging in productive exchanges of views. Through this kind of discourse, participants learn how to achieve an exchange of ideas that maintains respect and cooperation, consideration and tolerance toward others, as well as produces comprehensible expression and eloquent speech (Michaels et al., 2008).

The combination of argumentation principles and respectful discourse rules forms what is known as “Accountable Talk” (AT). During participation in Accountable Talk, participants engage in a continuous meta-cognitive process of constructing knowledge through higher-order thought processes characterized by elaborating ideas, critiquing, and drawing inferences (Chinn et al., 2001; Michaels et al., 2013). Through the process of accountable talk, individuals learn how elevated discourse enables the presentation of arguments in a clear, eloquent manner (Mercer, 2002; Reznitskaya et al., 2009), and the acquisition of refined speech conduct such as attentive and respectful listening to an interlocutor by for example, making appropriate eye contact. AT enhances learning by fostering critical thinking, active engagement, and evidence-based reasoning. It promotes crucial skills such as persuasive argumentation and respectful debate, thereby preparing individuals for constructive discourse and effective conflict resolution. Consequently, AT serves as a significant tool in nurturing collaborative, respectful, and effective communication skills (Michaels et al., 2008).

Accountable Talk aligns with Vygotsky’s theory which posits that learning occurs in a social context and within the range of a student’s ability when they are provided with appropriate guidance. As interactions occurring between teachers and students serve as a scaffold for knowledge acquisition (van de Pol et al., 2010, 2015), the AT framework is a hallmark of this type of scaffold (Michaels et al., 2008). Intended for elementary school students, Accountable Talk is based on models of social learning and cognitive theories and consists of an accountable commitment to three components: to a learning community (addressed by respectful discourse; Mercer, 2002) to rigorous thinking, and to accurate knowledge (the latter two consist of basic elements of argumentation: data, justification, and claim; Toulmin, 1958).

Accountability to a learning community refers to the participants’ responsibility for establishing respectful discourse, including creating a considerate, well-mannered tone or setting, actively listening to others, and constructing responses based on others’ arguments. It also includes expressing agreement with, opposition to, or requesting clarification if necessary (Mercer, 2002; Michaels et al., 2013). Moreover, this component asks participants to maintain discourse norms that demonstrate respect toward others, such as refraining from interrupting another speaker (Michaels et al., 2008).

The second component, accountability to rigorous thinking, refers to the responsibility in argumentation discourse to maintain standards of critical thinking that are necessary for providing explanations and presenting an idea in a comprehensive, rational manner. This component finds its expression through the utilization of argumentation principles such as the formation of productive arguments and the establishment of facts in a logical manner (Michaels et al., 2013). Through accountability to rigorous thinking, AT practitioners can challenge discourse partners’ statements as well as request (and/or provide) explanations for points made during interactions (Michaels et al., 2008).

The third component, accountability to accurate knowledge, refers to the participants’ responsibility in argumentation discourse to present their position or claims in a specific and accurate manner and to use a reliable fact-based information sources. Moreover, it involves presenting arguments in a manner that expresses an established and reasoned position and the flexibility to change one’s positions in accordance with what is said in the discussion. Responsibility for accurate knowledge means using the knowledge, background, and concepts learned to support their thought process (Michaels et al., 2013; Billings and Roberts, 2014).

Building on these principles, it is crucial to consider how discourse socialization practices and educational opportunities influence the implementation of Accountable Talk. Recent research has explored the interplay between discourse socialization practices and educational contexts, comparing children’s discursive practices at home and school (Heller, 2014; Koichu et al., 2022). This research suggests that the relationship between familial discourse socialization practices and the mastering of institutional communicative demands is not strictly linear or causal. Instead, a dynamic process unfolds in interactions, wherein teachers can establish a match between students’ discourse practices and institutional demands by making communicative investments. These investments are vital for students to perceive themselves as legitimate and competent members of the classroom discourse community.

The sense of participation in an institutional practice, as highlighted in the discourse socialization research, forms the foundation for utilizing interactions with teachers as external resources for acquiring knowledge and discourse competences (Quasthoff et al., 2022). When this sense of participation and membership is not established, students may withdraw from classroom discourse, leading to various interpretations such as refusal, lack of interest, or reluctance.

Several quantitative as well as qualitative research tools exist for examining respectful discourse and argumentation abilities in the classroom (Michaels et al., 2008; Suresh et al., 2019). One of the quantitative tools is a teacher questionnaire based on the Accountable Talk model (Michaels et al., 2008), and it separately examines each of the three core principles of considerate discourse, namely: accountability to a learning community, accountability to rigorous thinking, and accountability to accurate knowledge. The questionnaire, which examines both the performance of respectful discourse and evidence-based construction of arguments, is completed by teachers about their students (Schwarz and Baker, 2016; Glade, 2020). Another quantitative research tool is the Instructional Quality

Assessment (IQA) tool that is designed to assess students' skills during participation in whole group discussions and includes the assessment of the teacher's perspective on respectful discourse and their students' utilization of its principles. IQA data collection includes documentation of discourse procedures, researcher/teacher comments, and curation of the conversation data into a structured commentary according to the principles of respectful discourse (Boston, 2012).

In addition to those quantitative tools, currently available qualitative AT tools include classroom observation, grading and scoring of classroom conversations based on accountable talk elements, audio recordings (transcripts of discussions), documenting conversation procedures according to accountable talk structured checklists, and pre- and post-intervention interviews (Resnick and Hall, 1998; Matquma et al., 2006; Wolf et al., 2006). The existing research tools for assessing Accountable Talk skills are therefore rooted in observation and scoring of discourse conducted in the classroom. However, these tools have some shortcomings, such as significant costs (as raters have to be trained and compensated for the hours, sometimes days, of classroom observation involved) and interrater reliability (Suresh et al., 2019).

In contrast, self-report questionnaires allow for subjective self-assessment. The use of this test method is based on the premise that the participant is the prime resource and is most knowledgeable regarding their own conduct. The main advantages of self-report questionnaires are the relative simplicity of data collection, low administration costs, and the ability to increase respondent throughput within a limited time frame. This methodology also allows for the processing of results closer to the time of sampling.

As a review of the literature indicated there are no self-report questionnaires examining accountable talk abilities, we aimed to develop a questionnaire that examines argumentation abilities in the context of respectful discourse, and to provide initial validation for this self-report AT questionnaire (ATQ). The ATQ could then be utilized to assess improvement in argumentation and respectful discourse abilities, before and after an intervention program. The current study, therefore, aimed to construct a questionnaire that examined and validated the three main components of Accountable Talk: accountability to a learning community, accountability to rigorous thinking, and accountability to accurate knowledge. The questionnaire presented statements stratified into three sub-topics that describe different aspects of argumentation within the framework of respectful discourse.

Both quantitative and qualitative analysis was used for the development of the ATQ. Data obtained from the questionnaires completed by participants were analyzed quantitatively using Kaiser-Meyer-Olkin (KMO), Bartlett's test, Exploratory Factor Analysis (EFA), and test-retest reliability. For the qualitative analysis, data was obtained (recorded on camera) while participants conducted a group discussion on a topic provided by the experimenter. The observation data was transcribed and coded according to the items in the ATQ and then grouped into the three categories of the ATQ. Thus, the analyses are based on two levels: quantitative data obtained from the self-report questionnaire and a qualitative analysis based on transcribed group observations. Correlations between the two methods' data (quantitative data obtained from the ATQ and qualitative data obtained from the

observed group discussion) were computed to further validate the questionnaire.

## 2 Methods

### 2.1 Procedure

Following the approvals from the university's human research ethics committee and the chief scientist from the Ministry of Education, the study enrolled 100 students. To recruit participants, the researcher contacted two schools with similar socio-economic areas (cluster 5 out of 9, according to the Central Bureau of Statistics system) and obtained research approval from principals and teachers. All parents of participants were approached to provide signed informed consent to allow their children to participate in the study, and all student participants were explained the nature of the research and assented to their own participation. These students were organized into small groups and convened in a school setting outside their typical classrooms. Each student was tasked with completing the Accountable Talk Questionnaire (ATQ) in a single session, establishing the quantitative part of the data collection process. Half of the participants (50 students) were also randomly assigned to the qualitative part of the study. Thus, 50 students participated in both a quantitative and a qualitative study and 50 students participated only in the quantitative aspect of the study.

Following ATQ completion, qualitative data collection began in the same session. The participants were divided into seven groups (seven students in six groups and one group with eight students). A 45-min session was recorded on video that included a group discussion or group problem-solving about animal research. The experimenter positioned one video camera on a tripod close to each group to ensure that all interactions could be clearly recorded with minimal disturbance. To further avoid disturbing the interactions more than necessary, the researcher started and stopped the recordings at the beginning and end of each discussion, respectively. Transcriptions of these video recordings were subsequently analyzed and coded by two independent evaluators in accordance with the ATQ items. Three weeks following the initial session, 50 of the students were randomly selected and were asked to complete the ATQ again to assess test-retest reliability. This dual-mode approach of data collection provided a comprehensive understanding of AT abilities among students and facilitated the cross-validation of findings through a quantitative-qualitative correlation.

### 2.2 Participants

The sample comprised of 100 students: 52 males (52.0%) and 48 females (48.0%) who were native Hebrew speakers in age range of 11–12 years. The experiment was approved by the Education Ministry's chief scientist and the authors' university ethics committee. Data obtained from participants with a specific learning disability or ADHD (according to self-report) was excluded from the study. All students completed the ATQ. Among the 100 participants, 50 students were randomly selected to participate in group discussions. Data obtained from these

discussions was videotaped by the experimenter and was used for the qualitative analysis.

## 2.3 Data analyses

Several statistical analyses were conducted. First, after constructing the initial pool of statements for the ATQ, three judges were requested to review the existing research literature (Reznitskaya et al., 2001; Mercer, 2002; Michaels et al., 2008) and to assign each of the items into one of the three components of Accountable Talk. Next, Exploratory Factor Analysis (EFA) was conducted, limited to three orthogonal factors, only on statements that the judges had unanimously agreed regarding their assignment to one of the three ATQ components. A cohort sample size of 100 is considered large enough for EFA when there are no missing values (McNeish, 2017). Next, test-retest reliability was assessed with Pearson correlation analysis among a subset of 50 participants (mean age 11.5 years) at two separate times with a 3-week interval, and paired *t*-tests were used to evaluate the practice effect.

In addition, qualitative analysis was conducted among a subset of 50 participants. The qualitative data was gathered from the group discussions, processed through a coding scheme mirroring ATQ items. Each instance of accountable talk was identified and coded, with frequency counts calculated for each behavior. Mean responses to related items within each ATQ factor were calculated for each participant, resulting in scores that quantified their engagement with various aspects of Accountable Talk. Correlations were computed between ATQ scores and observed frequencies (for each factor), elucidating the relationship between self-reported accountable talk abilities and the observed behavior.

## 3 Results

### 3.1 Constructing the statements

Creation of the ATQ began with the identification of the three core components of accountable talk (Michaels et al., 2008). A set of statements reflecting these components were formulated for respondents to rate on a Likert scale. Thus, the initial questionnaire statement pool consisted of 20 statements (see Table 1) that were classified into three sub-topics addressing different aspects of argumentation in the context of respectful discourse. These sub-topics included: (1) responsibility to the community, containing seven statements, (2) responsibility to rigorous thinking, containing seven statements, and (3) Responsibility to accurate knowledge, containing six statements. In order to tailor the questionnaire to the specific objectives of the present study, we incorporated items from theories and models proposed by various researchers: Argument structure (Reznitskaya et al., 2001), respectful discourse (Mercer, 2008), and Accountable Talk (Michaels et al., 2008). Several statements in the ATQ were adapted from the Instructional Quality Assessment (IQA), a respected evaluation tool used for evaluating student performance. For instance, the ATQ statement, “In order to strengthen my claim, I use evidence and examples from the text,” was rephrased from the IQAs, “Were contributors asked to support their contributions

with evidence from the text?” The questionnaire asked participants to rate how much they agree or disagree with each statement, using a Likert scale ranging from 1–5 (5 = to a very large extent, 1 = not at all). The scoring of each item corresponded to the number chosen by the participant on the scale. Thus, each item on the questionnaire was scored 1–5 according to the participant’s rating: 1 point if participant selected “1” (not at all)- 5 points if participant selected “5” (to a very large extent). The ATQ exclusively employed positively aligned items, corresponding to the core components of accountable talk. This was designed to avert potential confusion and response bias that negatively phrased items can elicit, thereby enhancing the consistency, reliability, and validity of participant responses.

#### 3.1.1 Reliability between judges

Three judges who are experts in construction of measurements and evaluation tools were asked to review the existing research literature regarding Accountability Talk (Mercer, 2002; Michaels et al., 2008; Reznitskaya et al., 2009). The judges were then asked to examine the 20 proposed ATQ statements and to assign each to one of the three components of AT. The judges unanimously agreed on the assignment of 16 of the 20 statements. As no consensus was reached for the four following items—“I learn about different opinions on topics from group discussions,” “I speak in a focused and coherent manner,” “I would rather be quiet and not participate in a classroom discussion,” and “I phrase my arguments in accordance with the structure of the acceptable argument, which includes a positions, reasoning, counter-reasoning, and refutation”—those four statements were excluded from the questionnaire.

#### 3.1.2 Exploratory factor analysis

After removing the four items from the ATQ measure due to the lack of consensus between judges, EFA analysis limited to three orthogonal factors using Varimax rotation was conducted on the remaining 16 items. The items and factor loadings for the ATQ are presented in Table 1.

As can be seen in Table 1, the results of the EFA analysis indicated that the three orthogonal factors of the 16-item ATQ explained a total variance of 51.46%, with each factor explaining at least 14% of the additional variance. The 16 items were divided into three factors as follows: accountability to a learning community (e.g., “I maintain a pleasant tone after I hear an opposing opinion”), accountability to rigorous thinking (e.g., “During the discussion I state my opinion, whether it is in favor of others in the group or in opposition to them”), and accountability to accurate knowledge (e.g., “In order to strengthen my claim I use evidence and examples from the text”). For each factor, only items with factor loadings higher than 0.40 were chosen (Akpa et al., 2015), leading to the removal of an additional four items (statements). After their removal, another EFA analysis limited to three orthogonal factors using Varimax rotation was conducted on the remaining 12 items. We conducted Kaiser-Meyer-Olkin (KMO) and Bartlett’s tests of sphericity in order to examine whether it was plausible to conduct factor analysis for the 12 items of the ATQ questionnaire. The results indicated a medium-high KMO value of 0.76 and the

TABLE 1 Items and factor loadings for the 16-item ATQ.

Item	Descriptive		Factors		
	Mean	SD	1	2	3
1. In group discourse I listen to others speak.	4.35	0.90	0.79		
2. I hear out the other person without making comments.	3.74	1.10	0.82		
3. I maintain a pleasant tone after I hear an opposing opinion.	3.70	1.14	0.80		
4. I treat a team member with an opposing opinion with respect.	4.08	0.95	0.78		
5. When a team member speaks, I sit in a way that indicates my interest.	3.68	1.10	0.67		
6. In order to strengthen my claim, I use evidence and examples from the text.	3.79	1.20		0.74	
7. I choose my best reasoning to present to my opponents.	3.79	1.22		0.77	
8. When I try to strengthen my claim, I use prior knowledge.	3.94	0.89		0.73	
9. I prepare counter-reasoning to the reasoning of the party opposing my claim.	2.75	1.30		0.66	
10. During the discussion I state my opinion, whether it is in favor of others in the group or in opposition to them.	3.79	1.12			0.75
11. If I disagree with someone, I object to what they are saying and explain why.	3.89	1.11			0.83
12. I listen to what my team members are saying and respond to them.	3.87	1.06			0.60
13. I use “I” and not “you” messages.	3.38	1.00			<b>0.31</b>
14. If I have a hard time explaining myself during a discussion, I ask other team members for help.	2.70	1.53			<b>0.05</b>
15. My goal is to win the discussion.	2.94	1.43			<b>0.02</b>
16. It is important to me that the team members understand me.	4.54	0.81			<b>0.16</b>
<i>Eigenvalue</i>			4.03	2.72	1.48
$R^2$			20.94%	37.35%	51.46%
$\Delta R^2$			20.94%	16.41%	14.11%
Mean			3.91	3.57	3.59
Standard deviation (SD)			0.81	0.88	0.56

Only factor loadings above 0.40 were included in the medial version of the questionnaire (their coefficients are bolded).

Bartlett's test of sphericity indicated that the null hypothesis was rejected,  $\chi^2(66) = 469.13, p < 0.001$ . These results indicated that it was plausible to conduct factor analysis for the 12 items of the ATQ. The items and factor loadings for the ATQ measure are presented in Table 2.

As can be seen in Table 2, the EFA indicated that the three orthogonal factors of the 12-item ATQ explained a total variance of 64.98% (compared to 51.46% for the 16-item ATQ), with each factor explaining at least 18.3% of the additional variance. Thus, the EFA led to further item reduction by four due to low factor loadings, resulting in an ATQ of 12 items with an unequal distribution across factors (4, 5, 3). This distribution discrepancy, while potentially concerning, is not uncommon in scale development. The objective

of factor analysis is to identify clusters of items (factors) that measure the same underlying construct; it does not necessarily retain the initially proposed item allocations (Fabrigar et al., 1999; DeVellis and Thorpe, 2021). The factor loadings were higher than 0.58 and considered very high (Akpa et al., 2015). Moreover, the EFA indicated there were at least three items (statements) for each factor. This result takes into consideration the recommendation by Little et al. (1999) to only retain factors with at least three items.

### 3.1.3 Pearson correlations

After conducting the EFA, Pearson correlations were conducted to further establish the structural validity of the ATQ.



TABLE 2 Items and factor loadings for the 12-item ATQ.

Item	Descriptive		Factors		
	Mean	SD	1	2	3
1. I hear out the other person without making comments.	3.74	1.10	0.82		
2. I maintain a pleasant tone when I hear an opposing opinion.	3.70	1.14	0.81		
3. In group discourse I listen to others speak.	4.35	0.90	0.80		
4. I treat a team member with an opposing opinion with respect.	4.08	0.95	0.78		
5. When a team member speaks, I sit in a way that indicates my interest.	3.68	1.10	0.67		
6. In order to strengthen my claim I use evidence and examples from the text.	3.79	1.20		0.78	
7. I choose my best reasoning to present to my opponents.	3.79	1.22		0.77	
8. When I try to strengthen my claim, I use prior knowledge.	3.94	0.89		0.71	
9. I prepare counter reasoning for the reasoning of the party opposing my claim.	2.75	1.30		0.69	
10. If I disagree with someone, I object to what they are saying and explain why.	3.89	1.11			0.87
11. During the discussion I state my opinion, whether it is in favor of others in the group or in opposition to them.	3.79	1.12			0.87
12. I listen to what my team members are saying and respond to them.	3.87	1.06			0.58
<i>Eigenvalue</i>			3.85	2.65	1.29
<i>R</i> <sup>2</sup>			26.74%	46.67%	64.98%
$\Delta R^2$			26.74%	19.93%	18.31%
<i>Mean</i>			3.91	3.85	3.57
<i>Standard deviation (SD)</i>			0.81	0.92	0.88

Table 3 presents the Pearson correlation coefficients of the three components of Accountable Talk.

As can be seen in Table 3, the correlation pattern provided support for the structural validity of the ATQ (Table 2). All of the correlations were weak or mild, ranging from 0.08 to 0.43, supporting the distinction between different aspects of Accountable Talk.

### 3.1.4 Reliability of Cronbach’s Alpha (internal consistency)

The reliability of Cronbach’s Alpha (internal consistency) for the 12-item ATQ and for each of its three factors (components) was measured. The reliability of the 12-item ATQ was high,  $\alpha = 0.80$ . For the five items of the “accountability to a learning community” factor, the reliability was  $\alpha = 0.84$ . For the three items of the “accountability to rigorous thinking” factor, the reliability was  $\alpha = 0.78$ , and for the four items of the “accountability to accurate knowledge” factor, the reliability was  $\alpha = 0.75$ .

TABLE 3 Pearson correlation coefficients for the ATQ’s three components.

ATQ components	1	2	3
Accountability to a learning community	1	0.08	0.28**
Accountability to rigorous thinking		1	0.43***
Accountability to accurate knowledge			1

\*\*p < 0.01, \*\*\*p < 0.001.

### 3.1.5 Test-retest reliability

Test-retest reliability was assessed for all ATQ items, using a subset of 50 participants (mean age = 11.5 years, SD = 0.87). The participants were tested by the same examiner at

TABLE 4 Test-retest reliability and practice effect of ATQ ( $N = 50$ ).

T1		T2		Test-retest reliability	Score difference (retest minus test)	$t$ -test (test vs. retest)
Mean	SD	Mean	SD			
<b>Accountability to a learning community</b>						
4.01	0.62	4.02	0.62	0.98**	0.01**	$t(49) = 0.47, p = 0.642$
<b>Accountability rigorous thinking</b>						
3.97	0.50	3.99	0.50	0.99**	0.02**	$t(49) = 1.43, p = 0.159$
<b>Accountability to accurate knowledge</b>						
3.39	0.61	3.41	0.58	0.99**	0.02**	$t(49) = 0.00, p = 0.999$

\*\*Significant at the 0.01 level.

two time points separated by a 3-week interval. Table 4 includes number of test-retest indices, including test-retest reliability, practice effect, and  $t$ -test scores. Test-retest reliability was assessed with Pearson correlation coefficients. ATQ scores demonstrated excellent correlations ( $r > 0.98$ ). Practice effect was calculated as a mean of the difference in scores across time (Retest minus Test). As Table 4 shows, no significant differences were found between the two time points for all three factors of the ATQ. In addition, paired  $t$ -tests were used to assess the practice effect.

## 3.2 Qualitative analysis

Participants' discussions were observed and recorded for qualitative analysis. The purpose of this qualitative analysis was to validate the self-reported responses obtained from the ATQ. For that end, a qualitative analysis was conducted on a subset of 50 participants. The participants were divided into seven groups. Each group engaged in a 45-min discussion that was video-recorded. Two judges (doctoral students trained by the author to code the questionnaires) coded the transcribed videos. They were instructed to decide whether each statement among the 12-item of the ATQ was present or absent in the transcript. Each appearance of an item corresponding to one of the ATQ items received a score of 1. For example, the judges assigned a score of 1 if a participant presented a counterargument to a peer's supporting argument. In addition to the verbal communication of the participants, all relevant nonverbal communication (e.g., turning the body toward the speaker indicating listening or looking directly at the speaker) observed in the videos was also coded by the two judges. Inter-judge reliability for all 50 participants was found to be high (ICC = 0.997).

To examine whether any significant positive correlations existed between the three components of Accountable talk in the ATQ and in the recorded group discussion, Pearson correlation analyses were conducted. The observations corresponding to each ATQ factor were summed for each participant, resulting in cumulative scores that quantified their engagement with each factor of the Accountable Talk. These scores were subsequently utilized to conduct a correlation analysis between each of the ATQ's

TABLE 5 Pearson correlation coefficients between the three components of accountable talk in the ATQ and the recorded group discussion ( $N = 50, df = 48$ ).

Accountable talk components in ATQ	Accountability talk components in recorded group discussion		
	(1)	(2)	(3)
Accountability to a learning community	0.61***		
Accountability rigorous thinking		0.74***	
Accountability to accurate knowledge			0.42**

\*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

statements, investigating the relationship between the self-reported ATQ responses and the observed group discussions (see Table 5).

As can be seen in Table 5, significant positive medium-high correlation coefficients were found between the three components of Accountable Talk in the self-reported ATQ and the observed group discussion.

## 4 Discussion

The aim of the current study was to develop and provide initial validation for the ATQ. Our analyses confirmed the three main components of Accountable Talk, resulting in a compact 12-statement questionnaire divided into three sub-topics corresponding to the three main components of Accountable Talk within the framework of respectful discourse.

The contribution of this study is both theoretic and practical. The main theoretical contribution of this study is based on the identification of the same three Accountable Talk factors (components) for the ATQ that were previously identified by two independent research tools (IQA and a teacher questionnaire); this finding supports previous empirically and theoretically theorizing research in cognitive and social psychology about the nature of Accountable Talk that focusing on 3 major aspects to promote

students learning (Michaels et al., 2008) and is further evidence of its tripart factor structure.

The practical contributions of the present study include evidence that the ATQ can encompass the broad theory of argumentation principles within the framework of respectful discourse rules while utilizing only twelve statements. A measure that utilizes such a small number of items may avoid participant fatigue and serve as a tool that complements more involved existing AT assessment methodologies such as observation. Through the ATQ's relatively simple self-report methodology, its low administration costs, and its accessibility and convenience that may facilitate greater throughput of research participants over a limited timeframe, it may also enable completion of research within a shorter time frame and lower costs (by avoiding training observers and turning them into experts). By enable easier and faster monitoring and identification of individual and group abilities with regard to argumentation and respectful discourse, the ATQ is potentially well-positioned to evaluate the effectiveness of related school interventions and to examine what key factors influence student achievements in these areas.

The statements used in the current 12-item questionnaire were informed by previous research tools examining Accountable Talk, as well as theories and models developed by several researchers (Reznitskaya et al., 2001; Mercer, 2008; Michaels et al., 2008) and were further modified by the researchers. For example, one of the ATQ statements chosen to evaluate the "Accountability to a learning community" component is "I listen to what my team members are saying and respond to them." The corresponding statement that appears in the Instructional Quality Assessment (IQA) tool—designed to assess students' respectful discourse skills and utilization of the principles in group discussions—is "Did the speakers' contributions link to and build on each other?" Another example relates to the "Accountability to accurate knowledge" component of Accountable Talk. In the ATQ, we used the statement "In order to strengthen my claim I use evidence and examples from the text," while the statement that appears in the IQA toolkit was "Were contributors asked to support their contributions with evidence from the text?" Further similarities can be found in the statements examining the "Accountability to rigorous thinking" component. In the ATQ we used the statement "If I disagree with someone, I object to what they are saying and explain why," whereas the IQA toolkit's statement reads "Did contributors explain their thinking during the lesson?"

Similarly to the current ATQ, the Instructional Quality Assessment (IQA) tool is designed to assess the quality of classroom instruction on AT and how it manifests in the classroom, using the same three AT sub-topics (Resnick and Hall, 1998; Michaels et al., 2008). In the IQA toolkit study, Cronbach's alpha coefficients ranged from 0.74 to 0.92 and Spearman's correlation coefficients ranged from 0.62 to 0.83. Specifically, reliability of Cronbach's alpha for every factor of the IQA was as follows: "Accountability to a learning community"  $\alpha = 88$ ; "Accountability to rigorous thinking" was divided into "Asking for rigorous thinking"  $\alpha = 82$  and "Providing rigorous thinking"  $\alpha = 92$ , and "Accountability to accurate knowledge" was divided into "Asking for knowledge"  $\alpha = 74$  and "Providing knowledge"  $\alpha = 88$  (Resnick and Hall, 1998; Michaels et al., 2008). The present study also found good reliability for the 12 items of the ATQ ( $\alpha = 0.80$ ). For the five items of the

"Accountability to a learning community" factor the reliability was  $\alpha = 0.84$ , while the reliability of three items of the "Accountability to rigorous thinking" factor was  $\alpha = 0.78$  and the four items of the "Accountability to accurate knowledge" factor was  $\alpha = 0.75$ .

Further support for the structural validity of the 12-item ATQ was the correlation pattern that demonstrated weak or mild correlations between the three sub-topics, thus supporting the distinction between different aspects of Accountable Talk. Additionally, the results of the EFA indicated there were at least three items in each factor, and the findings of satisfactory internal reliability indicated that the items in each of the three factors had close associations as a group. The results of the EFA also indicated that the three orthogonal factors ensure that the ATQ features parsimony of scale items, thus minimizing respondent fatigue (Clark and Watson, 2016). In addition, the 12-item ATQ shows a satisfactory reliability, as measured by internal consistency and test-retest reliability. Analyses of test-retest data revealed excellent stability in scores of the three components, as assessed after a 3-week interval. The current study data provide evidence that this new measure reliably assesses three separate and independent qualities of Accountable Talk.

Observing the correlations between the self-reported data and the corresponding observational data indicate medium-high correlations (Table 5). The correlation between the self-reported data for the third factor (accountability to accurate knowledge) and its corresponded observational data shows a weaker correlation ( $r = 0.42$ ) compared to the first factor ( $r = 0.61$ ) and the second factor ( $r = 0.74$ ). However, this correlation is significant indicating that the third factor may capture the matter. A previous study that examined the relationships between self-reported and observational data related to management strategies in the classroom ( $N = 20$ ) also found medium-high correlations in the magnitude 0.48–0.53 (Clunies-Ross et al., 2008). We note that the observational data was acquired in one session while students were discussing only one topic (animals). Future studies using more topics for discussions is required to further validate the relationships between the self-reported factors and the observational data thus establishing the usability of the ATQ.

The present questionnaire was validated among children aged 11–12 years. At this age, the ability to process, criticize, and interpret information is considered sufficiently developed Kuhn et al. (1997). According to Piaget (1997), a child of this age might be able to think in an abstract logical way that allows thinking about hypothetical problems, making abstract logical "moves," developing theories, and understanding complex causality. This is reflected in the ATQ statements. For instance, statement "I prepare counter-reasoning to the reasoning of the party opposing my claim" presupposes the ability to make abstract logical moves. The statement "If I disagree with someone, I object to what they are saying and explain why" is requires the ability to think hypothetically and to ground one's arguments with supporting evidence. Indeed, meta-cognitive thinking ability is critically necessary for participation in Accountable Talk, during which the participant is required to distinguish between thinking about knowledge and knowledge itself, between basic understanding of mental states and the understanding that the other's behavior may be influenced by different and/or opposing desires and beliefs. The ATQ statements "In group discourse I listen to others



speak” and “I treat a team member with an opposing opinion with respect” demonstrate that the questionnaire distinguishes between knowledge of the other’s mental state and the other’s opposing views.

One of the limitations of the current study lies in the self-report format of the Accountable Talk Questionnaire. Some researchers point out self-image biases (Paulhus and Reid, 1991), impression management, and competency and self-interest issues might undermine the validity of self-report questionnaire results. On the other hand, other scholars argue that these biases do not affect the predictive validity of such questionnaires. Indeed, several researchers have demonstrated that impression management and self-deception did not create spurious effects nor did they attenuate the predictive validities (Hough et al., 1990; Barrick and Mount, 1996). Furthermore, good correlations were found between the ATQ items identified by judges who observed the recorded group discussion and those same items on the self-reported ATQ, lending support to the reliability of the self-report questionnaire.

Another research limitation might relate to the questionnaire having been validated only with children aged 11–12 years. Although the questionnaire has been content-validated, and can be presumed suitable for older ages, future research calls for the validation of the questionnaire among those of older ages. In addition, the questionnaire contains sub-scales with a small number of items: five statements in the “accountability to a learning community” sub-scale, four items in the “accountability to accurate knowledge” sub-scale, and three items in the “accountability to rigorous thinking” sub-scale. Often, a small number of items can lower the value of alpha, but in the present study a good alpha reliability value was obtained. Nevertheless, although this number is sufficient for building a sub-scale, further research is recommended to add more items to increase reliability. While the number of items for each factor are unequal and somewhat small, these limitations are offset by the questionnaire’s overall psychometric properties. The ATQ has proven to be a reliable and valid tool, as indicated by its performance in both quantitative analyses and its correlation with observed behaviors. In future research, however, it may be beneficial to further refine and potentially expand the items for each factor to continue improving the ATQ’s robustness.

Furthermore, the absence of direct comparisons with other self-report questionnaires that assess Accountable Talk (AT) abilities is another limitation. However, to the best of our knowledge, there are currently no other self-report questionnaires that specifically evaluate AT abilities. Thus, the ATQ, while filling a notable gap, presented a challenge for the traditional validation process that typically involves correlating a new tool with established measures assessing the same construct. In lieu of a conventional approach, a theoretical link was carefully established between the ATQ and the constructs it aims to measure, drawing upon an exhaustive review of AT literature. Further strengthening the validation, parallels were drawn between the ATQ and the Instructional Quality Assessment (IQA), a recognized tool in the field. The IQA, while not a self-report measure, evaluates similar AT constructs within classroom settings. This connection with the IQA offers indirect substantiation of the relevance and applicability of the ATQ. Furthermore, the IQA contributed to the creation of the ATQ beyond theoretical parallels. Elements of the IQA were adapted and

integrated into the ATQ, with some statements’ wording modified to fit the requirements of a self-report questionnaire. This process, acknowledging the influence and validity of the IQA in assessing AT constructs, further solidified the indirect validation and the potential of the ATQ in this uncharted territory.

One of the future applications for this questionnaire pending further validation might be examining performance of populations that show difficulty in argumentation within the framework of respectful discourse rules, such as participants with attention deficit hyperactivity disorder (ADHD). Moreover, the questionnaire may be used as a tool to examine the performance of populations challenged in terms of social and verbal skills such as autism, which is known to suffer from communication impairments. Another limitation concerns the age range use in the current study. Future research should refine and expand the questionnaire’s items and validate it with different age groups.

## 5 Conclusion

The current study developed and validated the Accountable Talk Questionnaire (ATQ), a concise 12-item questionnaire that assesses the three main components of Accountable Talk within the framework of respectful discourse. The questionnaire showed good reliability and structural validity, with weak correlations between sub-topics, supporting the distinction between different aspects of Accountable Talk. The ATQ’s self-report methodology, low administration costs, and accessibility make it a valuable tool for assessing argumentation and respectful discourse skills, complementing existing assessment methodologies. Thus, the ATQ has the potential to assess Accountable Talk skills rapidly and economically among students. In addition, the ATQ has potential applications in evaluating the effectiveness of school interventions aiming to enhance AT skills. By doing so, the ATQ may assist research and learning communities to enhance deliberative democracy—emphasizing consultation, persuasion, and discussion—as a means of reaching rational decisions and thus better preparing students for a morally responsible adulthood.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by the Chief Scientist from the Ministry of Education. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants’ legal guardians/next of kin.

## Author contributions

AKT collected the data, analyzed the data, and wrote the paper. NM designed the study, analyzed the data, conceptualized the

study, and supervised the study. All authors contributed to the article and approved the submitted version.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Akpa, O. M., Elijah, A. B., and Baiyewu, O. (2015). The adolescents' psychosocial functioning inventory (APFI): scale development and initial validation using exploratory and confirmatory factor analysis. *Afr. J. Psychol. Study Soc. Issues* 18, 1–21.
- Barrick, M. R., and Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *J. Appl. Psychol.* 81, 261–272. doi: 10.1037/0021-9010.81.3.261
- Billings, L., and Roberts, T. (2014). From mindless to meaningful. *Educ. Leadership* 72, 60–65.
- Boston, M. D. (2012). Assessing the quality of mathematics instruction. *Elementar. Sch. J.* 113, 76–104. doi: 10.1086/666387
- Bova, A., and Arcidiacono, F. (2017). "Interpersonal dynamics within argumentative interactions: an introduction," in *Interpersonal Argumentation in Educational and Professional Contexts*, eds F. Arcidiacono, and A. Bova (New York, NY: Springer), 18–22.
- Bova, A., and Arcidiacono, F. (2023). The integration of different approaches to study the interpersonal dynamics of argumentation. *Learn. Cult. Soc. Inter.* 41, 100735. doi: 10.1016/j.lcsi.2023.100735
- Chinn, C., Anderson, R. C., and Waggoner, M. (2001). Patterns of discourse during two kinds of literature discussion. *Reading Res. Q.* 36, 378–411. doi: 10.1598/RRQ.36.4.3
- Clark, L. A., and Watson, D. (2016). "Constructing validity: basic issues in objective scale development" in *Methodological Issues and Strategies in Clinical Research*, ed A. E. Kazdin (New York, NY: American Psychological Association), 187–203.
- Clunies-Ross, P., Little, E., and Kienhuis, M. (2008). Self-reported and actual use of proactive and reactive classroom management strategies and their relationship with teacher stress and student behaviour. *Educ. Psychol. Int. J. Exp. Educ. Psychol.* 28, 693–710. doi: 10.1080/01443410802206700
- DeVellis, R. F., and Thorpe, C. T. (2021). *Scale Development: Theory and Applications*. London: Sage publications.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., and Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychol. Methods* 4, 272. doi: 10.1037/1082-989X.4.3.272
- Glade, J. A. (2020). *The Influence of Number Talks on the Use of Accountable Talk*. [Master thesis, Northwestern College, Iowa] Orange City, IA: Northwestern College.
- Hasnunidah, N., Susilo, H., Irawati, M., and Suwono, H. (2020). The contribution of argumentation and critical thinking skills on students' concept understanding in different learning models. *J. Univ. Teach. Learn. Prac.* 17, 6. doi: 10.53761/1.17.1.6
- Heller, V. (2014). Discursive practices in family dinner talk and classroom discourse: a contextual comparison. *Learn. Cult. Soc. Int.* 3, 134–145. doi: 10.1016/j.lcsi.2014.02.001
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., and McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *J. Appl. Psychol.* 75, 581–595. doi: 10.1037/0021-9010.75.5.581
- Koichu, B., Schwarz, B. B., Heyd-Metzuyanim, E., Tabach, M., and Yarden, A. (2022). Design practices and principles for promoting dialogic argumentation via interdisciplinarity. *Learn. Cult. Soc. Int.* 37, 100657. doi: 10.1016/j.lcsi.2022.100657
- Kuhn, D., Shaw, V., and Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognit. Instr.* 15, 287–315. doi: 10.1207/s1532690xci1503\_1
- Kuhn, D., and Udell, W. (2003). The development of argument skills. *Child Dev.* 74, 1245–1260. doi: 10.1111/1467-8624.00605
- Lazarou, D., Erduran, S., and Sutherland, R. (2017). Argumentation in science education as an evolving concept: Following the object of activity. *Learn. Cult. Soc. Int.* 14, 51–66. doi: 10.1016/j.lcsi.2017.05.003
- Little, T. D., Lindenberger, U., and Nesselrode, J. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: when "good" indicators are bad and "bad" indicators are good. *Psychol. Methods* 4, 192–211. doi: 10.1037/1082-989X.4.2.192
- Matquma, L. C., Slater, S. C., Wolf, M. K., Crosson, A., Levison, A., Peterson, M., et al. (2006). *Using the Instructional Quality Assessment Toolkit to Investigate the Quality of Reading Comprehension Assignments (CSE Technical Report 669)*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), 14–30.
- McNeish, D. (2017). Exploratory factor analysis with small samples and missing data. *J. Pers. Assessment* 99, 637–652. doi: 10.1080/00223891.2016.1252382
- Mercer, N. (2002). "Developing dialogues," in *Learning for Life in the 21st Century: Sociocultural Perspectives on the Future of Education*, eds G. Wells and G. Claxton (London: Blackwell), 141–153.
- Mercer, N. (2008). The seeds of time: why classroom dialogue needs a temporal analysis. *J. Learn. Sci.* 17, 33–59. doi: 10.1080/1058400701793182
- Michaels, S., Hall, M. W., and Resnick, L. B. (2013). *Accountable Talk Sourcebook: For Classroom Conversation That Works*. Pittsburgh, PA: University of Pittsburgh.
- Michaels, S., O'Connor, C., and Resnick, L. B. (2008). Deliberative discourse idealized and realized: accountable talk in the classroom and in civic life. *Stu. Philos. Educ.* 27, 283–297. doi: 10.1007/s11217-007-9071-1
- Paulhus, D. L., and Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *J. Pers. Soc. Psychol.* 60, 307–317. doi: 10.1037/0022-3514.60.2.307
- Piaget, J. (1997). *The Principles of Genetic Epistemology, Vol. 7*. London: Psychology Press.
- Quasthoff, U., Heller, V., Prediger, S., and Erath, K. (2022). Learning in and through classroom interaction: on the convergence of language and content learning opportunities in subject-matter learning. *Eur. J. Appl. Ling.* 10, 57–85. doi: 10.1515/eujal-2020-0015
- Resnick, L. B., and Hall, M. W. (1998). Learning organizations for sustainable education reform. *Daedalus* 127, 89–118.
- Reznitskaya, A., and Anderson, R., NcNulen, B. (2001). Influence of Oral discussion on written argument. *Discourse Proc.* 32, 155–175. doi: 10.1207/S15326950DP3202andamp;3\_04
- Reznitskaya, A., Kuo, L. J., Clark, A. M., Miller, B., Jadallah, M., Anderson, R. C., et al. (2009). Collaborative reasoning: a dialogic approach to group discussions. *Cambridge J. Educ.* 39, 29–48. doi: 10.1080/03057640802701952
- Schwarz, B. B., and Baker, M. J. (2016). *Dialogue, Argumentation and Education: History, Theory and Practice*. Cambridge: Cambridge University Press.
- Suresh, A., Sumner, T., Jacobs, J., Folland, B., and Ward, W. (2019). Automating analysis and feedback to improve mathematics teachers' classroom discourse. *Pro. AAAI Conf. Artif. Int.* 33, 9721–9728. doi: 10.1609/aaai.v33i01.33019721
- Toulmin, S. E. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.
- van de Pol, J., Volman, M., and Beishuizen, J. (2010). Scaffolding in teacher-student interaction: a decade of research. *Educ. Psychol. Rev.* 22, 271–296. doi: 10.1007/s10648-010-9127-6
- van de Pol, J., Volman, M., Oort, F., and Beishuizen, J. (2015). The effects of scaffolding in the classroom: support contingency and student independent working time in relation to student achievement, task effort and appreciation of support. *Instruc. Sci.* 43, 615–641. doi: 10.1007/s11251-015-9351-z
- Webb, N. M., Franke, M. L., Ing, M., Turrou, A. C., Johnson, N. C., Zimmerman, J., et al. (2019). Teacher practices that promote productive dialogue and learning in mathematics classrooms. *Int. J. Educ. Res.* 97, 176–186. doi: 10.1016/j.ijer.2017.07.009
- Wolf, M. K., Crosson, A. C., and Resnick, L. B. (2006). *Accountable Talk in Reading Comprehension Instruction. CSE Technical Report 670*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), 48–82.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.