

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Salokivi, Maija; Salanterä, Sanna; Saarikallio, Suvi; Ala-Ruona, Esa

Title: Validation of item pool for early adolescents' emotional skills assessment in music therapy

Year: 2024

Version: Published version

Copyright: © 2023 the Authors

Rights: CC BY 4.0

Rights url: <https://creativecommons.org/licenses/by/4.0/>

Please cite the original version:

Salokivi, M., Salanterä, S., Saarikallio, S., & Ala-Ruona, E. (2024). Validation of item pool for early adolescents' emotional skills assessment in music therapy. *Nordic Journal of Music Therapy*, 33(2), 108-121. <https://doi.org/10.1080/08098131.2023.2245433>



32 (5) 2023

Validation of item pool for early adolescents' emotional skills assessment in music therapy

Maija Salokivi, Sanna Salanterä, Suvi Saarikallio & Esa Ala-Ruona

To cite this article: Maija Salokivi, Sanna Salanterä, Suvi Saarikallio & Esa Ala-Ruona (2023): Validation of item pool for early adolescents' emotional skills assessment in music therapy, Nordic Journal of Music Therapy, DOI: [10.1080/08098131.2023.2245433](https://doi.org/10.1080/08098131.2023.2245433)

To link to this article: <https://doi.org/10.1080/08098131.2023.2245433>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 04 Sep 2023.



[Submit your article to this journal](#)




[View related articles](#)



[View Crossmark data](#)

Validation of item pool for early adolescents' emotional skills assessment in music therapy

Maija Salokivi ^a, Sanna Salanterä^b, Suvi Saarikallio^c and Esa Ala-Ruona^d

^aDepartment of Music, Art and Culture Studies, University of Jyväskylä, Jyväskylä, Finland; ^bClinical Nursing Science, University of Turku, Turku, Finland; ^cMusic Education, University of Jyväskylä, Jyväskylä, Finland; ^dMusic Therapy, University of Jyväskylä, Jyväskylä, Finland

ABSTRACT

Introduction: Evaluating the validity of the content is an essential step in developing an assessment tool, including an analysis of the quality of the items within the tool. This study describes the content validation of items in the early adolescents' emotional skills assessment tool in a music therapy context.

Methods: Content validity was evaluated based on relevance scores provided by two expert panels. Psychometric scores were obtained by calculating item-specific content validity index (I-CVI), scale-specific content validity index (S-CVI) and modified kappa score. In addition, the coverage and understandability of the items were evaluated.

Results: The validation process identified 60 valid items distributed across six components of emotional skills: expressing, monitoring, identifying, understanding, regulating and the ability to use emotional information. Item I-CVI scores ranged from 0.80 to 1.00, the scale content validity index (S-CVI) was 0.95, the modified kappa score ranged from 0.65 to 1.00, item coverage at scale level was 1.00, and item understandability was 0.92.

Discussion: The items developed in the study have high validity and are scientifically grounded. The items can be a first step towards a validated assessment tool to evaluate emotional skills in early adolescents. The added value of this study is that the set of items developed is the first to cover all the components of emotional skills identified in the literature. Therefore, music therapists can use the items to observe in more detail the different dimensions of emotional skills in early adolescents.


ARTICLE HISTORY Received 9 September 2022; Accepted 18 July 2023

KEYWORDS Emotional skills; early adolescents; content validity; item analysis; music therapy assessment

Introduction

Early adolescents (around 11–13) undergo many biological, externally visible changes, but their inner emotional development is also progressing. These changes in emotional development include greater emotional reactivity; an increased ability to reflect on

CONTACT Maija Salokivi  majja.salokivi@gmail.com 

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/08098131.2023.2245433>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

emotions and to evaluate the acceptability and expression of emotions; and new strategies to manage emotions (Davey et al., 2008; Steinberg, 2005). Unfortunately, many early adolescents have a range of mental health conditions that complicate their years of adolescence. Mental health conditions such as depression, anxiety, behavioural disorders, and developmental disability are the leading causes of illness and disability among young people (10–19 years old). Unfortunately, half of these conditions begin already by the time a young person reaches 14 years of age (World Health Organization, 2021). Children and young people with mental health conditions face a wide range of risks, including poor school performance, poor subjective emotional well-being, behavioural problems and adverse life events such as dropping out of school (Aviles et al., 2006; McLeod et al., 2012; Tempelaar et al., 2014). In addition, these mental health conditions are often accompanied by deficits in emotional and interpersonal skills (Gonçalves et al., 2019; Parker et al., 2006).

In Finland, where the authors of this study come from, early adolescents with mental health conditions are the largest group of music therapy clients. Early adolescents account for 60% of all children (0–15 years old) using music therapy, and 85% of the early adolescents in music therapy have mental health problems (Social Insurance Institution of Finland, 2017). The significant size of the client group, the importance of emotional skills in the lives of children and adolescents, and the practical benefits of music therapy with children and adolescents with mental health conditions (Geipel et al., 2018; Gold et al., 2004, 2007) are a justifiable reason to seek a validated assessment tool for the progress of early adolescents' emotional skills in a music therapy context. In addition, the requirement for evidence-based practice, the expectations of funders, and the desire of music therapists to communicate the impact and effectiveness of their work make the development of an assessment tool meaningful to the field (Cripps et al., 2016).

There is currently no validated and reliable assessment tool in music therapy for assessing emotional skills in early adolescents (Cripps et al., 2016). Moreover, existing assessment tools only consider a limited number of emotional skills and their different components, not all of which have been defined in the scientific literature (Salokivi et al., 2022). Existing music therapy assessment tools that are available for children and adolescents assess different aspects of emotional skills, such as emotional expression (Langan, 2009; Mackeith et al., 2011), emotional responsiveness (Layman et al., 2002), emotional constriction (Wells, 1988), social-emotional functioning or behaviour (Douglas, 2006; Goodman, 1989), affect range (Loewy, 2000), and differentiation, expression, regulation and self-awareness (Baxter et al., 2007).

This study uses the term “early adolescents’ emotional skills” even though the English term “emotion” does not have an unambiguous meaning, and there is no scientific consensus on the definition of the term (Frijda, 2016; Izard, 2010; Lakoff, 2016). The term “emotion” is a multi-component and multi-level concept (Zachar, 2010), and it can be viewed as an everyday concept or as a scientific concept (Widen & Russell, 2010). The difference between an everyday and scientific concept can be examined by separating the descriptive and prescriptive definitions. The descriptive definition refers to the definition of a word in everyday life, and the prescriptive definition refers to the scientific concept of a word, which includes a set of events that scientific theory aims to explain (Widen & Russell, 2010). In this study, we use the descriptive definition of The APA Dictionary of Psychology of the word “emotion”: “a complex pattern of reactions, including experiential, behavioural and physiological

elements, through which an individual attempts to process a personally significant issue or event” (American Psychological Association, 2023).

Validating the items of the assessment tool through an operationalisation process

The validity of an assessment tool means the extent to which it measures the characteristics of the issue being studied (Devon et al., 2007). The validity of the assessment tool includes content, construct and criterion validity (Lynn, 1986). Content validity is a prerequisite for other types of validity and should therefore be a priority for assessment developers (Zamanzadeh et al., 2015). Content validity assessment examines how well a sample of items captures the operational definition of a concept (Polit & Beck, 2006) and to what extent the items adequately represent the content domain (Carmines & Zeller, 1979; Yaghmaie, 2003).

This study presents the content validation process of an item pool used to assess emotional skills in early adolescents with mental health conditions in the context of music therapy. The research questions for this study are: What kind of item pool will be generated based on the results of previous studies (Salokivi et al., 2022, 2023)? Does the item pool have adequate content validity?

The progression of content validation of the item pool is presented through a five-step operationalisation process: (1) developing a theoretical definition; (2) specifying variables derived from the theoretical definition; (3) identifying observable indicators; (4) selecting means of measuring the indicators; and (5) evaluating the adequacy of the resulting operational definition (Waltz et al., 2016). The starting point (step 1) for the validation process is carefully determining the concept (Salkind, 2012). We reported on the process of concept definition in our previous study (Salokivi et al., 2022), in which the components of emotional skills in early adolescents were theoretically defined through a scoping review and concept analysis. The second (2) and third (3) steps of the operationalisation we reported in the study by Salokivi et al. (2023). In those steps, we applied the components of early adolescents’ emotional skills to the music therapy clinical practice. We used focus group interviews and content analysis to explore how music therapists worked with the different components of emotional skills in early adolescents, what methods they use, and where they can see the progress of the emotional skills during the therapy process. This knowledge was a foundation for the next steps presented in this study: (4) selecting means of measuring the indicators and (5) evaluating the adequacy of the resulting operational definition. In step four, the item pool is formulated, and in step five, the adequacy of the created items is evaluated based on their psychometric values. [Figure 1](#) shows the validation process for the items.

Methods

The item pool initially included 62 items, which fell within the six primary components of the conceptual model of emotional skills: expressing emotions; monitoring emotions; identifying emotions; understanding emotions; regulating emotions; and the ability to use emotional information for self-management and establishing social relationships (Salokivi et al., 2022). Each component contained several items. The number of items in the subcomponents ranged from four to 14. The first author of this study formulated the item pool based on the results of two previous studies

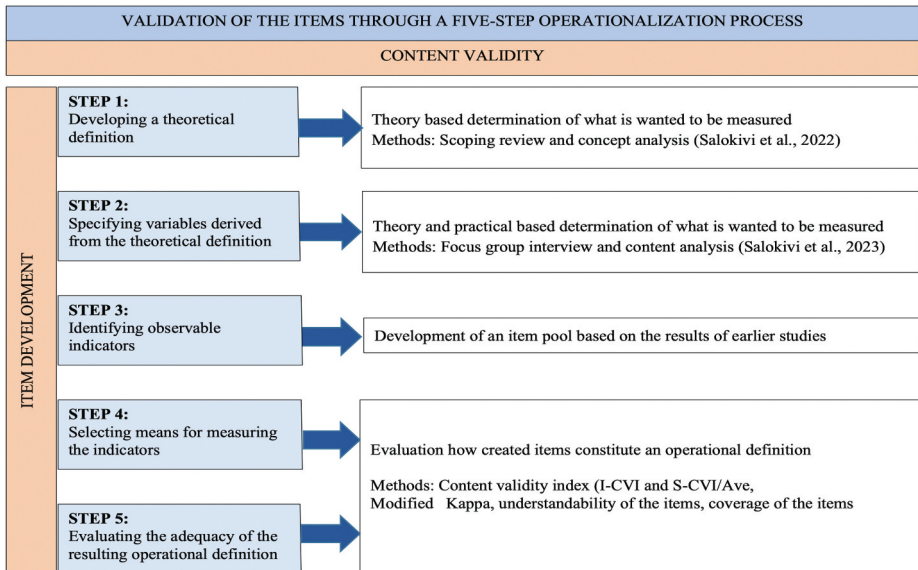


Figure 1. Validation of the items through a five-step operationalisation process.

(Salokivi et al., 2022, 2023). The senior researcher, also the second author of this article, commented on the items and suggested changes. After the items were re-edited, a final set of items was formed, and the content validity of these items was analysed.

Analyzing items

The relevance of the items was studied based on the answers of expert panels through the Content Validity Index (CVI) and Modified Kappa. Psychometric scores were also calculated for the understandability of the items and coverage of the scale. In addition, qualitative written comments were collected from expert panellists to help further develop the questions where necessary.

Participants of the expert panels

Two groups of panellists were involved in the first round of the item evaluation for relevance, understandability, and coverage. One group comprised health researchers ($n=8$), and the other comprised music therapy clinicians ($n=8$). Health researchers represented theoretical expertise, and music therapists represented clinical expertise. The health researchers were doctoral researchers who volunteered to participate. Instrument development is a part of their education, and they are familiar with evaluating items, especially from the perspective of written form and understandability. Music therapists were volunteer clinicians who participated in an earlier phase of the research project and consented to participate in this phase of the study. They were specialists in the content and context. The responses of the music therapists and health researchers were

examined separately for possible divergences. The research team were second-round panellists ($n = 3$) and evaluated the understandability of the reformulated items when needed. The panellists were senior researchers with instrument development and music therapy expertise.

Content validity index

Content validity was calculated at the item level using the Item Content Validity Index (I-CVI) and the scale level with the Scale Content Validity Index (S-CVI) (Lynn, 1986). In this study, the I-CVI and S-CVI were examined as follows: The panels of experts were asked to rate each item on the scale regarding its relevance to the underlying concept (Polit & Beck, 2006). The developed questionnaire had a 4-point ordinal scale (1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, and 4 = highly relevant) (Lynn, 1986; Waltz & Bausell, 1981).

The I-CVI was calculated as follows: the number of experts who gave the question a score of 3 or 4 divided by the total number of experts (Polit & Beck, 2006). The I-CVI had to be at least 0.78 for an item to be accepted. The recommended threshold for a panel of six to ten experts is 0.78 (Lynn, 1986). If the I-CVI of an item was less than 0.78 when the panellist groups' responses were calculated separately, both groups' average I-CVI determined whether the item was accepted.

The S-CVI in this study was an S-CVI/Ave (average), which is defined as the average I-CVI value at the scale level and is a recommendation for the S-CVI calculation when there are several members in the expert panel (Polit & Beck, 2006; Waltz & Bausell, 1981). S-CVI/Ave was calculated by summing the I-CVI values of the 3- and 4-classified items, divided by the total number of items. The value had to be at least .90 for the scale to be accepted (Polit & Beck, 2006).

Modified kappa

The modified kappa (k^*) is the consensus among evaluators that the item is relevant and considers the possibility of chance agreement (Polit et al., 2007). The kappa values were evaluated to determine whether they were fair, good, or excellent (Cicchetti & Sparrow, 1981; Fleiss, 1971). Polit et al. provided scale developers with a ready-made table for estimating k^* values based on the number of experts and their agreement. Scale developers can compare their survey I-CVI values with this table without having to calculate the modified kappa. This study used a pre-established table, which is presented in Table 1.

However, if the researcher wants to calculate the k^* value, the first information needed is the subject's random agreement. The random agreement can be examined using a binomial random variable formula that calculates how often a given event occurs in a fixed number of trials or attempts (Polit et al., 2007; Ross, 2017):

$$P_c = \left[\frac{N!}{A!(N-A)!} \right] .5^N$$

P_c refers to the probability of chance agreement, N is the number of experts, and A is the number agreeing on good relevance. Sign !, after N and A , means that the number is a factorial function; in other words, one must multiply all whole numbers from the chosen number to 1 (for example, $3! = 3 \times 2 \times 1 = 6$). The second step is

Table 1. Evaluation of I-CVIs with different numbers of experts and agreement.

(1)	(2)	(3)a	(4)b	(5)c	(6)d
Number of Experts	Number Giving Rating of 3 or 4	1-CVI	<i>P</i> _c	<i>k</i> [*]	Evaluation
3	3	1.00	.125	1.00	Excellent
3	2	.67	.375	.47	Fair
4	4	1.00	.063	1.00	Excellent
4	3	.75	.25	.67	Good
5	5	1.00	.041	1.00	Excellent
5	4	.80	.156	.76	Excellent
6	6	1.00	.016	1.00	Excellent
6	5	.83	.094	.81	Excellent
6	4	.67	.234	.57	Fair
7	7	1.00	.008	1.00	Excellent
7	6	.86	.055	.85	Excellent
7	5	.71	.164	.65	Good
8	8	1.00	.004	1.00	Excellent
8	7	.88	.031	.88	Excellent
8	6	.75	.109	.72	Good
9	9	1.00	.002	1.00	Excellent
9	8	.89	.014	.89	Excellent
9	7	.78	.070	.76	Excellent

^a1-CVI, item level content validity index.

^b*P*_c (probability of a chance occurrence) was computed using the formula for a binomial random variable, with one specific outcome: $P_c = [N!/A!(N - A)!] \cdot .5^N$ where *N* = number of experts and *A* = number agreeing on good relevance.

^c*k*^{*} = kappa designating agreement on relevance: $k^* = (1 - \text{CVI} - p_c) / (1 - p_c)$.

^dEvaluation criteria for kappa, using guidelines described in Cicchetti and Sparrow (1981) and Fleiss (1971): Fair = *k* of .40 to .59; Good = *k* of .60–.74; and Excellent = *k* > .74.

From "Polit et al. (2007), Is the CVI an acceptable indicator of content validity? Appraisal and recommendations". Copyright 2007 by John Wiley & Sons.

calculating the I-CVI (the proportion of agreements on relevance) and the probability of chance agreement (Polit et al., 2007).

Understandability and the coverage of the items

The panel of experts was asked to rate each item for understandability. The scale used was the same as in the relevance evaluation: a 4-point ordinal scale (1 = not clear, 2 = somewhat clear, 3 = quite clear, and 4 = highly clear). If the understandability of the item was < 0.78 (the same threshold value as in the relevance calculation), the item was reformulated and developed further with the second round's smaller group of senior researchers (*n* = 3) until the understandability of the item was at an acceptable level. Because of the small size of this group, a universal agreement was required before the item was accepted. In this context, the universal agreement meant that all panellists gave a score of 3 or 4 for the item.

The expert panellists had the opportunity to write a comment on each item if they wanted to clarify why the item was difficult to understand or if they wanted to give their recommendations for rewording the item. These open-ended qualitative comments were used to develop better items if the items still needed to receive an appropriate validity score in the first round.

The panels of experts were also asked to evaluate the items' coverage of the concept (emotional skills of early adolescents). The panellists received a short semi-structured questionnaire about coverage. The scale for the evaluation was a 4-point ordinal scale

(1 = strongly disagree, 2 = somewhat disagree, 3 = somewhat agree, and 4 = strongly agree) (David, 1992). They could also comment if they wanted to add or remove items or comment on something else concerning the item.

Translation of the items

The panellists were Finnish, and the original items were in Finnish. The final item pool's Finnish items were translated into English using DeepL Translator, a neural machine translator that uses an artificial neural network (DeepL GmbH, 2017). After translating the AI, a native speaker of both Finnish and English translated the English items back into Finnish. The result of the back-translation was compared to the original Finnish items, and if some of the items needed to be reworded to improve understandability, changes were made. However, the back-translation process did not significantly alter the English translation of the Deep L translator. The meanings of the items remained the same. The AI translation used “emotion” and “feeling” as parallel terms in English, while only one term was used in the original Finnish version. In the translated English items, the term “emotion” was retained. Thus, the final English language set of items was created.

Ethical considerations

The study received ethical approval from the Human Sciences Ethics Committee of the University of Jyväskylä (Number: 746/13.00.04.00/2020). Participants in expert panels received the study's fact sheet and data protection notification. The panellists were aware that they had the right to refuse to participate in the study or to discontinue participation if they wished, without penalty. Data were saved and stored in a secure data environment. The participants in the expert panel were not identifiable in the research report.

Results

The results of this study describe the content validity of the items in the item pool to be used to assess emotional skills in early adolescents. The items in the item pool were categorised according to the components of the emotional skills of early adolescents (Salokivi et al., 2022). The emotional skills components were: expressing, monitoring, identifying, understanding, and regulating emotions; and the ability to use emotional information both on an individual level to develop and establish positive self-management and on a social level to develop and establish positive relationships.

Content validity

Item content validity index, scale content validity index and modified kappa

After averaging the I-CVI values of the original 62 items between the groups, only one item had an I-CVI value below 0.78. This item was deleted before the full-scale content validity index, and modified kappa was calculated. When calculating the average of both panellists' groups, item-specific I-CVI values ranged from 0.80 to 1.00 for 61 items, which is an excellent result (Cicchetti & Sparrow, 1981; Fleiss, 1971). The more detailed results of the expert panels in terms of I-CVI scores for all 62 items were as

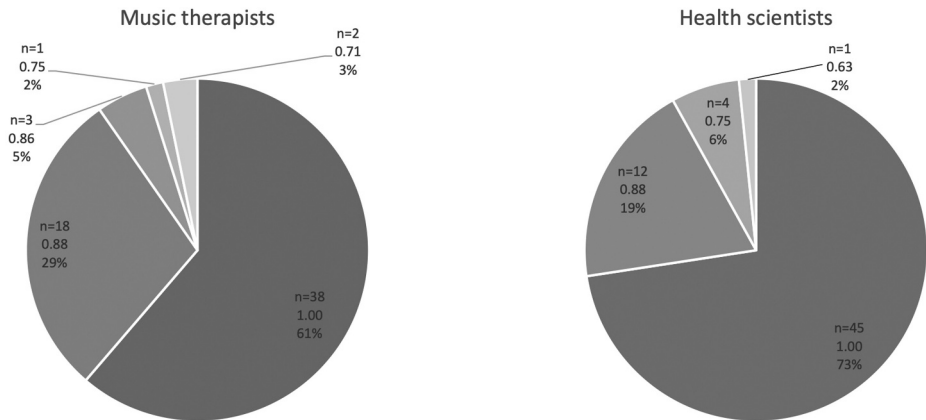


Figure 2. Percentage distribution of relevance scores for the I-CVI expert panels.

follows: Music therapists: 38 items 1.00, 18 items 0.88, three items 0.86, one item 0.75 and two items 0.71 and health researchers: 45 items 1.00, 12 items 0.88, four items 0.75 and one item 0.63. A detailed item-specific table of I-CVI relevance values and modified kappa values can be found in Table S1. The appendix contains all values examined for the items and is included as online supplemental material due to its large size. [Figure 2](#) shows the I-CVI distributions of the expert panels as a percentage.

The set of items in the following stages of the study consisted of 61 items. The average of the S-CVI/Ave values of the panellist groups for the relevance of the 61 items was 0.95, which is an excellent result (Cicchetti & Sparrow, 1981; Fleiss, 1971). The modified kappa evaluation of the 61 items ranged from 0.65 to 1.00 (good to excellent), indicating an acceptable level of agreement (Cicchetti & Sparrow, 1981; Fleiss, 1971). Most items had excellent modified kappa values between 0.88–1.00 from both panellist groups. The expert group opinions were as follows. Music therapists: 38 items 1.00, 18 items 0.88, three items 0.85, and two items 0.65; health researchers: 45 items 1.00, 12 items 0.88, four items 0.72.

Understandability and the coverage of the items

The Content Validity Index for the understandability of the items and the overall scale was counted for relevance in the same way as the I-CVI and S-CVI/Ave. Understandability I-CVI values for the 61 tentative items yielded results between 0.63–1.00. Expert groups' opinions were as follows: music therapists: 27 items 1.00, 24 items 0.88, one item 0.86, eight items 0.75, and one item 0.63; health researchers: 35 items 1.00, 20 items 0.88, two items 0.75, and four items 0.63. [Figure 3](#) presents the distribution of the I-CVI understandability results of the expert panels in percentages.

The understandability of the fifteen items was below the acceptable value of 0.78. When averaging the I-CVI scores of both expert groups for the understandability of these fifteen questions, the scores for the five items still ranged from 0.63 to 0.75 and were under the acceptable value of 0.78. Therefore, these five items went to a small group of senior experts for further review. The five items were reformulated to improve their understandability until a consensus score of 1.00, a universal agreement, was reached. One item did not reach a universal agreement on understandability and

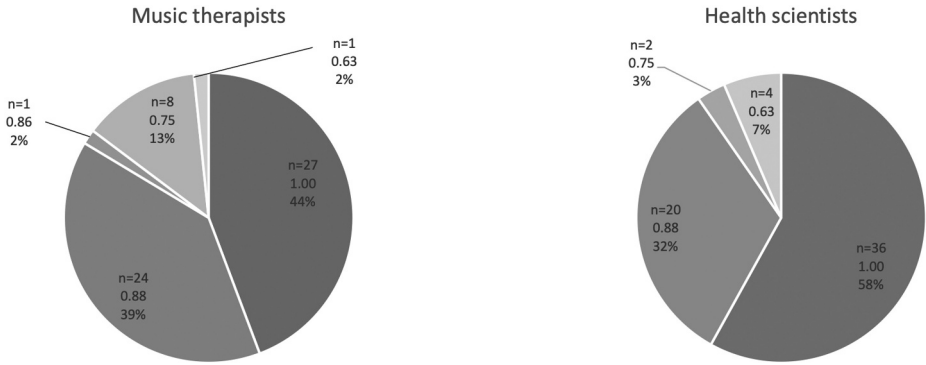


Figure 3. Percentage distribution of understandability scores for the I-CVI expert panels.

Table 2. Reformulation of the four items.

Original item	Reformulated item
3. Expressing emotions non-verbally by oneself	3. Expressing feelings by oneself when being alone (writing, making music, painting)
4. Personal contents when listening to or playing music	4. Expresses personal contents in music listened to or played
36. Musical expression is structured	36. Can play/sing in rhythm
37. Other expression is structured	37. Verbal expression is understandable

was removed from the set of items at this stage. Detailed information on all the I-CVI understandability values of the items is provided in Table S1. Table 2 presents the reformulation of the remaining four items that reached a universal agreement (1.00) among all senior experts.

After the understandability assessments described in the two previous paragraphs and the necessary rewording of the four items by a small group of panellists, the final set of 60 items with I-CVI values between 0.86 and 1.00 were selected. These final items are presented in Table S2. Next, the scale-level understandability of the final set of 60 items was counted by calculating the average understandability of the items. The average for understandability was 0.92.

In addition, the coverage of the concept was calculated, and the I-CVI was used to assess coverage. Both music therapists and health researchers covered the items with a value of 1.00. The experts did not recommend adding any items to the pool. At the end of the validation evaluation, the item pool comprised 60 validated items.

Discussion

Based on this study, the 60 items selected for the final item pool were valid and acceptable regarding the psychometric results. The content validity process identified 60 items that fell within six dimensions: expressing emotions (eight items); monitoring emotions (four items); identifying emotions (six items); understanding emotions (four items); regulating emotions (14 items); and an ability to use emotional information both on self-management (14 items) and establishing social relationships (10 items). All 60 items included in the final item pool received excellent scores for relevance (0.80–1.00), understandability (0.86–1.00), scale-

level understandability (0.92) and scale content validity (S-CVI) (0.95). In addition, the modified kappa score for chance agreement ranged from good to excellent (0.65–1.00), and the scale coverage was rated high (1.00). Based on the results, the item pool generated by the study has a high level of validity, is scientifically grounded, and is helpful for further research.

There were no significant differences between the responses of the music therapists and health researchers. The similarity in the responses of the expert groups and the strong content validity scores of items may reflect the impact of the careful development process that preceded this study. Our two previous studies laid the groundwork for this study and the developed item pool. In the first study, we focused on developing a theoretical definition (Salokivi et al., 2022). In our second study, we examined the applicability of the theoretical definition to music therapy practice (Salokivi et al., 2023).

The results of this study may provide valuable insights for validation research on assessing emotional skills in early adolescents undergoing music therapy. Few existing music therapy assessment tools, potentially also applicable to assess emotional skills in early adolescents, report how their validity has been assessed (Cripps et al., 2016). This makes it difficult to evaluate the psychometric quality of assessment tools or the items they contain. This study describes in detail the content validity analysis of the item pool of the assessment tool. In addition to the reported content validity assessment, the added value of this study is that the item pool developed in this study is the first to cover the different components of emotional skills as described in the research literature. In this study, we have used a specific description of the components of emotional skills that we developed earlier (Salokivi et al., 2022). The description helps us better conceptualise all the components of emotional skills.

Furthermore, we have examined the phenomenon of emotional skills in practice in the context of music therapy (Salokivi et al., 2023), created practice-based items, and assessed the content validity of these items. The result is a first step towards a validated assessment tool. The items used in the final assessment tool will still be further refined and developed. However, the current item pool comprises a detailed description, based on current knowledge, of the phenomenon of emotional skills in music therapy when working with early adolescents and, as such, offers new insights for music therapists and music therapy researchers.

The linguistic formulation of the items developed in this study does not necessarily correspond to the formulation used in the final assessment tool. The linguistic format may change when the items are tested in practical music therapy work. In addition, the perspective of the items in this study has been kept at a general level without taking a definitive position on who will ultimately use the items (therapist or client etc.). Testing the items in further studies in music therapy practice will provide the information needed to determine what size of the item set is adequate and what kind of linguistic formulation of the items will be most suitable. Therefore, the item pool developed is only an intermediate step towards the items used in the final assessment tool.

Waldon et al. (2018) state, “Without professionally developed and standardised assessment tools, the field of music therapy is less robust and less equipped to meet the demands of a respected health care profession” (p. 42). The results of this study are a preliminary step towards a validated assessment tool to assess the emotional skills of early adolescents with mental health conditions. Furthermore, a validated item pool

may already be helpful to music therapy clinicians in their practice. They can use these items to learn and observe different components of emotional skills in early adolescents and use this information to improve the well-being of their young clients.

Strengths and weaknesses

As Spiro et al. (2020) noted, the feasibility and psychometric properties of an assessment tool are critical issues in the selection and usefulness of that tool. From this perspective, the developed item pool for emotional skills of early adolescents with mental health conditions has strengths and weaknesses. Based on this study, the 60 remaining items after the psychometric evaluation have strong content validity, and the item development has been appropriate and careful. However, the 60-item assessment tool in music therapy is still too long, and its practical feasibility is not necessarily excellent. Additionally, the items in the item pool have not yet been tested in a clinical context, and it is possible that, for example, some of the items are not relevant in a therapy practice or some items overlap. It should also be noted that the items presented to the expert panels do not represent all possible items related to emotional skills, and some essential items may be missing. This limitation was minimised by conducting two previous studies to ensure that the concept and its operationalisation were comprehensive and scientifically grounded. In addition, the experts could suggest additional items for the set of items, which also helped to minimise the possible absence of essential items.

Future research

The validated item pool can be used as a foundation for the assessment tool, which could then be tested and developed further to obtain psychometrically valid, clinically relevant, and feasible tools. Content validity evaluation of the assessment tool should continue with reliability evaluation (through internal consistency and test-retest), construct validity (through factor analysis), and criterion-related validity (an estimate of the extent to which a measure agrees with a gold standard) (Grant & Davis, 1997).

Conclusions

This study presents the first set of 60 items with high content validity to assess the emotional skills of early adolescents with mental health conditions in the context of music therapy. The study describes how content validity was assessed regarding the item content validity index, scale content validity index/average, modified kappa scores, and item understandability and coverage. This set of items is a preliminary step towards a validated tool for assessing emotional skills in early adolescents, which still needs to be created and does not yet exist in music therapy. A validated assessment tool will strengthen music therapists' ability to assess emotional skills development more accurately during the music therapy process. In addition, a validated assessment tool will also help therapists better articulate music therapy's impact on clients, guardians, and therapy funders.

Disclosure statement

The authors report no conflict of interest.

Funding

No funding was received for this study.

Notes on contributors

Maija Salokivi, MA, is a music therapist and a doctoral researcher at the University of Jyväskylä, Finland. She has worked with children and adolescents with psychiatric and neurological problems for several years. Her PhD dissertation aims to develop an assessment instrument for the emotional skills of early adolescents in the context of music therapy. Her research interests include assessment, assessment development, emotional skills, and children and adolescents.

Sanna Salanterä, PhD, RN, Professor of Clinical Nursing Science, Vice Dean of the Faculty of Medicine at the University of Turku and Nurse Director at Turku University Hospital. She specialised in clinical nursing science, nursing and patient decision-making, innovative technology in nursing, pain nursing, health service research, children's nursing and empowering patient education. She leads the research programs "Health producing basic care with smart technology" and "Health producing health promotion with smart technology". Both projects consist of several sub-projects.

Suvi Saarikallio, PhD, Professor of Music Education, Docent of Psychology at the University of Jyväskylä, Finland. She is a music psychology research team member and widely conducts research and teaching in music psychology, music education, and music therapy. She is currently President of the Finnish Society for Music Education (FiSME) and the European Society for the Cognitive Sciences of Music (ESCOM) Secretary General.

Esa Ala-Ruona, PhD, Associate Professor, music therapist and psychotherapist working at the Music Therapy Clinic for Research and Training at the University of Jyväskylä, Finland. His research interests are music therapy assessment and evaluation, studying musical interaction and clinical processes in improvisational psychodynamic music therapy, and the progress and outcomes of rehabilitation of stroke patients in active music therapy.

ORCID

Maija Salokivi  <http://orcid.org/0000-0003-0886-3532>

References

- American Psychological Association. (2023). *Dictionary of psychology*. Emotion. Retrieved January 2023, from <https://dictionary.apa.org/emotion>
- Aviles, A. M., Anderson, T. R., & Davila, E. R. (2006). Child and adolescent social-emotional development within the context of school. *Child and Adolescent Mental Health, 11*(1), 32–39. <https://doi.org/10.1111/j.1475-3588.2005.00365.x>
- Baxter, H., Berghofer, J., MacEwan, L., Nelson, J., Peters, K., & Roberts, P. (2007). *The individualized music therapy assessment profile*. Jessica Kingsley Publishers.
- Carmines, E., & Zeller, R. (1979). *Reliability and validity assessment*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412985642>
- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*(2), 127–137.

- Cripps, C., Tsiris, G., & Spiro, N. (Eds.). (2016). *Outcome measures in music therapy: A resource developed by the Nordoff Robbins research team*. Nordoff Robbins. <https://www.nordoff-robbins.org.uk/research-projects/outcome-measures-in-music-therapy/>
- Davey, C. G., Yücel, M., & Allen, N. B. (2008). The emergence of depression in adolescence: Development of the prefrontal cortex and the representation of reward. *Neuroscience and Biobehavioral Reviews*, 32(1), 1–19. <https://doi.org/10.1016/j.neubiorev.2007.04.016>
- David, L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research*, 5(2), 194–197. [https://doi.org/10.1016/S0897-1897\(05\)80008-4](https://doi.org/10.1016/S0897-1897(05)80008-4)
- DeepL GmbH. (2017). DeepL translator (4.5.542015)[Computer software]. DeepL SE. <https://www.deepl.com/translator>
- Devon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship*, 39(2), 155–164. <https://doi.org/10.1111/j.1547-5069.2007.00161.x>
- Douglas, E. (2006). The development of a music therapy assessment tool for hospitalized children. *Music Therapy Perspectives*, 24(2), 73–79. <https://doi.org/10.1093/mtp/24.2.73>
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382. <https://doi.org/10.1037/h0031619>
- Frijda, N. H. (2016). The evolutionary emergence of what we call “emotions”. *Cognition and Emotion*, 30(4), 609–620. <https://doi.org/10.1080/02699931.2016.1145106>
- Geipel, J., Koenig, J., Hillecke, T. K., Resch, F., & Kaess, M. (2018). Music-based interventions to reduce internalizing symptoms in children and adolescents: A meta-analysis. *Journal of Affective Disorders*, 225, 647–656. <https://doi.org/10.1016/j.jad.2017.08.035>
- Gold, C., Voracek, M., & Wigram, T. (2004). Effects of music therapy for children and adolescents with psychopathology: A meta-analysis. *Journal of Child Psychology and Psychiatry*, 45(6), 1054–1063. <https://doi.org/10.1111/j.1469-7610.2004.t01-1-00298.x>
- Gold, C., Wigram, T., & Voracek, M. (2007). Effectiveness of music therapy for children and adolescents with psychopathology: A quasi-experimental study. *Psychotherapy Research*, 17(3), 289–296. <https://doi.org/10.1080/10503300600607886>
- Gonçalves, S. F., Chaplin, T. M., Turpyn, C. C., Niehaus, C. E., Curby, T. W., Sinha, R., & Ansell, E. B. (2019). Difficulties in emotion regulation predict depressive symptom trajectory from early to middle adolescence. *Child Psychiatry and Human Development*, 50(4), 618–630. <https://doi.org/10.1007/s10578-019-00867-8>
- Goodman, K. (1989). Music therapy assessment of emotionally disturbed children. *The Arts in Psychotherapy*, 16(3), 179–192. [https://doi.org/10.1016/0197-4556\(89\)90021-X](https://doi.org/10.1016/0197-4556(89)90021-X)
- Grant, J. S., & Davis, L. L. (1997). Focus on quantitative methods: Selection and use of content experts for instrument development. *Research in Nursing & Health*, 20(3), 269–274. [https://doi.org/10.1002/\(sici\)1098-240x\(199706\)20:3<269:aid-nur9>3.3.co;2-3](https://doi.org/10.1002/(sici)1098-240x(199706)20:3<269:aid-nur9>3.3.co;2-3)
- Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review*, 2(4), 363–370. <https://doi.org/10.1177/1754073910374661>
- Lakoff, G. (2016). Language and emotion. *Emotion Review*, 8(3), 269–273. <https://doi.org/10.1177/1754073915595097>
- Langan, D. (2009). A music therapy assessment tool for special education: Incorporating education outcomes. *Australian Journal of Music Therapy*, 20, 78–98. <https://link.gale.com/apps/doc/A352616845/AONE?u=anon~9bb1b5f9&sid=googleScholar&xid=996e9fb0>
- Layman, D. L., Hussey, D. L., & Laing, S. J. (2002). Music therapy assessment for severely emotionally disturbed children: A pilot study. *Journal of Music Therapy*, 39(3), 164–187. <https://doi.org/10.1093/jmt/39.3.164>
- Loewy, J. (2000). Music psychotherapy assessment. *Music Therapy Perspectives*, 18(1), 47–58. <https://doi.org/10.1093/mtp/18.1.47>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–386. <https://doi.org/10.1097/00006199-198611000-00017>
- Mackeith, J., Burns, S., & Lindeck, J. (2011). The music therapy star: The outcomes star for children in music therapy. *Triangle Consulting Social Enterprise*, 2–18.
- McLeod, J. D., Uemura, R., & Rohman, S. (2012). Adolescent mental health, behavior problems and academic achievement. *Journal of Health and Social Behavior*, 53(4), 482–497. <https://doi.org/10.1177/0022146512462888>

- Parker, J. G., Rubin, K. H., Erath, S. A., Wojslawowicz, J. C., & Buskirk, A. A. (2006). Peer relationships, child development, and adjustment: A developmental psychopathology perspective. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Theory and method* (pp. 419–493). John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470939383.ch12>
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing and Health*, 29(5), 489–497. <https://doi.org/10.1002/nur.20147>
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Focus on research methods: Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing and Health*, 30(4), 469–467. <https://doi.org/10.1002/nur.20199>
- Ross, S. M. (2017). Discrete random variables. In S. M. Ross (Ed.), *Introductory statistics* (pp. 203–262). Academic Press. <https://doi.org/10.1016/b978-0-12-804317-2.00005-9>
- Salkind, N. (2012). Content validity. In N. Salkind (Ed.), *Encyclopedia of research design*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412961288>
- Salokivi, M., Salanterä, S., & Ala-Ruona, E. (2022). Scoping review and concept analysis of early adolescents' emotional skills: Towards development of a music therapy assessment tool. *Nordic Journal of Music Therapy*, 31(1), 63–88. <https://doi.org/10.1080/08098131.2021.1903977>
- Salokivi, M., Salanterä, S., Saarikallio, S., & Ala-Ruona, E. (2023). Promoting emotional skills in early adolescents with mental health conditions in music therapy – A content analysis of focus group interviews. *Nordic Journal of Music Therapy* 32 (5), 423–444. <https://doi.org/10.1080/08098131.2023.2169336>
- Social Insurance Institution of Finland. (2017). *Kelan kuntoutustilasto*. <https://helda.helsinki.fi/handle/10138/234527>
- Spiro, N., Tsiris, G., & Cripps, C. (2020). Sounds good, but . . . what is it? An introduction to outcome measurement from a music therapy perspective. *Approaches: An Interdisciplinary Journal of Music Therapy*, 12(1), 8–29. <https://approaches.gr/spiro-a20180627/>
- Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences*, 9(2), 69–74. <https://doi.org/10.1016/j.tics.2004.12.005>
- Tempelaar, W. M., Otjes, C. P., Bun, C. J., Plevier, C. M., van Gastel, W. A., MacCabe, J. H., Kahn, R. S., & Boks, M. P. M. (2014). Delayed school progression and mental health problems in adolescence: A population-based study in 10,803 adolescents. *BMC Psychiatry*, 14(244). <https://doi.org/10.1186/s12888-014-0244-5>
- Waldon, E. G., Lindahl Jacobsen, S., & Gattino, G. (2018). Assessment in music therapy: Psychometric and theoretical considerations. In S. L. Jacobsen, E. G. Waldon, & G. Gattino (Eds.), *Music therapy assessment: Theory, research and application* (pp. 47–80). Jessica Kingsley Publishers.
- Waltz, C. F., & Bausell, R. B. (1981). *Nursing research: Design, statistics, and computer analysis*. F.A. Davis.
- Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2016). *Measurement in nursing and health research*. Springer Publishing Company. <https://doi.org/10.1891/9780826170620>
- Wells, N. (1988). An individual music therapy assessment procedure for emotionally disturbed young adolescents. *The Arts in Psychotherapy*, 15(1), 47–54. [https://doi.org/10.1016/0197-4556\(88\)90051-2](https://doi.org/10.1016/0197-4556(88)90051-2)
- Widen, S. C., & Russell, J. A. (2010). Descriptive and prescriptive definitions of emotion. *Emotion Review*, 2(4), 377–378. <https://doi.org/10.1177/1754073910374667>
- World Health Organization. (2021, August 9). *Adolescent mental health*. <https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health>
- Yaghmaie, F. (2003). Content validity and its estimation. *Journal of Medical Education*, 3(1), 25–27. <https://brieflands.com/articles/jme-105015.html>
- Zachar, P. (2010). Has there been conceptual progress in the science of emotion? *Emotion Review*, 2 (4), 381–382. <https://doi.org/10.1177/1754073910374668>
- Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and implementation content validity study: Development of an instrument for measuring patient-centred communication. *Tabriz University of Medical Sciences*, 4(2), 165–178. <https://doi.org/10.15171/jcs.2015.017>