Creative Thinking in Music: Its Nature and Assessment Through Musical Exploratory Behaviors

Baptiste Barbot
Yale University

Todd Lubart
Université Paris Descartes

Creative thinking in music has received only limited attention in the psychology of art and creativity, yet it appears to be one of the most important issues in the field of music education. As creative thinking in music exists in the general population and increasing evidence suggests the positive effects of active engagement with music, it seems that research on this topic offers promising implications beyond musical education itself. However, there is a lack of suitable measures of creative ability in music for individuals without prior musical training, and many important facets, such as the original use of sound material, are often disregarded in existing assessments. To fill these conceptual and empirical gaps—a prerequisite to any research on musical creativity—we present a new multimethod framework for its evaluation: the Musical Expression Test (MET). This method combines a systematic observational approach with a focus on musical exploratory behaviors and a product-based assessment of musical pieces resulting from musical activities, involving a sound-production set and a computer-based recording system. A study with the MET on a sample of adolescents with and without musical training is presented and provides the first empirical evidence of the MET’s reliability, convergence between behavioral and product-based assessment, and suitability for individuals without particular musical skills. A typology of product-based “creative styles” and their behavioral correlates is presented. The potential uses and implications of the MET for the future of musical creativity assessment and research are discussed.

Keywords: creativity, music, divergent thinking, exploratory behavior, problem finding

The lives of eminent creative musicians, such as Bach, Mozart, and Beethoven, have received attention for centuries. However, the empirical study of creative thinking in music is relatively recent. Some literature exists in various fields, including psychology, cognitive science, and artificial intelligence, but most of the research on this topic has been conducted in the field of music education. Indeed, music education generally considers creativity as a fundamentally important dimension that should be promoted and integrated in educational programs (e.g., Webster, 1990). Consequently, research on musical creativity has mainly focused on children and adults with musical training (as a target of music education research). Nevertheless, musical creativity exists presumably in the general population as (a) it is accepted that creativity is a psychological dimension normally distributed (Guilford, 1950) and (b) humans show “spontaneous” musical activity, including the perception and comprehension of music early in life (e.g., Bigand & Poulin-Charronnat, 2006; Kogan, 1994). The social representation of musical creativity, long associated with eminence and giftedness (Adorno, 1976), together with conceptual debates on its definition and measurement, has delayed the systematic study of musical creativity in the general population.

However, as new music technologies become increasingly popular, many people are now involved in amateur creative-musical production (e.g., Mellor, 2007; Mellor, 2008; Partti & Karlsen, 2010; Ward, 2009). An increasing number of studies show the positive effects of active engagement with music on intellectual, social, and personal development (Hallam, 2010). Thus, the study of musical creativity—as a general psychological dimension—is a contemporary research topic with important implications beyond musical education concerns. Yet, a fundamental challenge to advance the research in musical creativity is its assessment. After surveying the state of the art on defining and measuring musical creativity, we present the Musical Expression Test (MET), a new multimethod technique to evaluate creative thinking in music through an innovative behavioral measurement of exploratory musical activity and a classic product-based assessment method. Finally, we present an empirical study using the MET with a sample of adolescents, to estimate the MET’s psychometric qualities, and confirm its suitability for the study of musical creativity in the general population.

Defining and Measuring Musical Creativity

The definition of musical creativity is a research topic in itself. However, there is consensus on a general definition of creativity,
which applies to music and other domains of creative work. This definition proposes that creativity is the ability to produce something that is both new/original and valuable/appropriate to the task or the domain (e.g., Amabile, 1996; Barron, 1988; Sternberg & Lubart, 1995). Whereas the originality component can be understood in a straightforward manner, some explanations of “appropriateness” in the music domain have been proposed. As “music” refers to organized sounds distinct from noise, most often “appropriateness” refers to the notion of “structure” (e.g., Auh, 2000), as a creative musical piece is not “random.” Similarly, appropriateness may refer to the notion of a “plan,” as originality without intent or a plan does not necessarily make a creative product (e.g., Hickey & Webster, 2001). However, the notion of appropriateness in a musical piece does not suggest that the production must be technically mastered or “musically correct.” Therefore, Webster (1990, 1992) deplored that general aptitudes in music (e.g., rhythmic regularity) are often confounded with creative aptitude in music. Webster (1990) also recommended the study of “creative thinking in music” rather than “musical creativity” to demystify the concept while focusing on the creative process and its role in music. Creative thinking in music is defined as a dynamic mental process that alternates between divergent and convergent thinking, moving in stages over time, enabled by internal musical skills and external conditions, and resulting in a final musical product that is new for the creator (Webster, 1990). The final musical product might be a composition or an improvisation, but the creative thinking process may also be involved in other musical activities such as musical performance (e.g., Clarke, 2005) or listening (Dunn, 1997).

Along with these conceptual distinctions, there are many methodological and technical challenges related to the development of valid assessments of musical creativity (Webster, 1992). A common way to deal with these issues consists of proposing both a definition of creativity and a method for its assessment (Pachet, 2006). Consequently, there are many instruments, procedures, and methods that have been developed to assess musical creative processes, products, or persons. A systematic review of these tools is not the purpose here (for this see, e.g., Hickey, 2002; Leung, Wan, & Lee, 2009; Nielsen, 2011; Richardson, 1983; Webster, 1992). Rather, we present an overview of the main measurement approaches—including musical divergent thinking tests, compositional creativity assessment, and product-based assessment—to discuss the critical needs for future measures of musical creativity.

**Measures of Musical Divergent Thinking**

Musical divergent thinking tests measure the thinking process of generating many responses to a single musical stimulus or problem (e.g., Baltzer, 1988; Gorder, 1980; Richardson, 1983; Webster, 1994), by means of musical fluency, flexibility, and originality scores. Musical fluency corresponds to the “productivity” of the musical ideas in response to a given stimulus or problem (e.g., to propose as many melodies as possible starting from three imposed notes). It is typically measured through the duration of various improvisations or instrumental explorations (e.g., Webster, 1994) or by the number of musical events produced in response to a task (e.g., Gorder, 1980). Musical flexibility refers to the variety of the musical content (e.g., Gorder, 1980) or the number of categories in which the responses can be classified. For instance, Webster’s (1994) Measure of Creative Thinking in Music II (MCTM-II) measures flexibility by categorizing the changes within three musical parameters: tempo (fast/slow), dynamics (loud/soft), and pitch (high/low). Finally, originality refers to the statistical frequency of musical responses with regard to the corpus of responses observed in a reference sample (e.g., Gorder, 1980; Webster, 1994).

Major issues with the measurement of musical divergent thinking include (a) the multidimensional nature of the construct (i.e., there are many parameters leading to the production of flexible and original musical ideas including, e.g., pitch, tempo, dynamics, choice of sound materials, or the rhythmic and melodic content) and (b) measures of flexibility and originality often involve a transcription of the responses using the traditional music notation (usually based on audio or video recordings). This transcription requires accomplished musicians to capture the “imperfections” and parameters (e.g., dynamics, tone) that the musical answer involves and which may represent the “true originality” of the musical response. Finally, (c) these measures are developed by music educators for music students, often limiting their applicability to this specific, academically trained population.

**Compositional Creativity Assessment**

Three main approaches are typically used to measure creative thinking in composing music: think-aloud or retrospective semistructured interviews (e.g., Barrett, 1996; Bennett, 1976), analyses of musical flow recorded at various stages of the compositional process (e.g., Folkestad, Hargreaves, & Lindström, 1998; Hickey, 1997; Seddon & O’Neill, 2003), or systematic observation of subjects composing music with codification of the musical activity (e.g., Kratus, 1989). The major issue with the study of the compositional process in music relates to the segmentation, definition, and identification of the processes, “stages,” or actions, which take place during the creative process. For example, Kratus (1994) focused on compositional activity through the processes of exploration, development, repetition, and silence. Even though the creative process conceptually varies based on the musical activity (e.g., composition, improvisation, performance), it is not clear whether this difference can be perceived objectively (e.g., Lehmann & Kopiez, 2010).

**Product-based Assessment**

Product-based assessment is a common way to evaluate the result of creative thinking in music. It generally involves the achievement of a musical piece (composition or improvisation) based on initial task constraints. Completed in standardized conditions, the resulting products are often scored using the Consensual Assessment Technique (CAT; Amabile, 1982, 1996). This technique involves domain-appropriate judges (e.g., music teachers, composers) who evaluate independently the musical products using a Likert scale on several criteria such as originality, musical structure, expressivity (e.g., Auh & Walker, 1999), and/or other “technical” criteria such as tonal and rhythmic coherence (e.g., Hickey, 2001; Kratus, 1994). The consensus between judges is
estimated statistically using correlational or Rasch measurement approach (Leung et al., 2009) to ensure interrater agreement and the reliability of the composite scores derived for each musical product. Typically, the agreement estimated in studies of musical creativity using the CAT is acceptable for various types of products, criteria, and experimental conditions (e.g., Auh, 1997; Hickey, 2001; Priest, 2006).

Perspectives for the Measurement of Creative Thinking in Music

Although many research efforts have concentrated on the development of broad and valid measures of creative thinking in music (e.g., Webster’s MCTM-II described above), there is a need to assess musical creativity with alternative methods (Hickey, 2002). In addition to their conceptual and methodological limits, existing measures of musical creativity often involve tasks that require a minimum of domain-specific knowledge (i.e., basic mastery of a musical instrument and/or of musical notation), which limits their applicability to samples of individuals with no prior music training. In contrast, popular creative thinking assessment tools in other domains (e.g., Torrance, 1966) do not require such basic skills, resulting in imbalanced advances in creativity research across domains.

In addition, important dimensions of musical creativity are often disregarded in the assessment literature. Particularly, the ability to produce unusual sonorities based on the use of new materials or the unexpected use of familiar instruments and objects is an important facet of musical creativity that could be even more important than musicality (e.g., Leman, 1999). In fact, this ability has historically led to major field innovation, as evidenced by findings in music archaeology. Many examples of the first vestige of human musical creativity, such as flutes and whistles, derived from objects of the environment (e.g., bones of birds) were dated as early as the early upper Palaeolithic age (e.g., d’Errico et al., 2003). More recently, the unexpected use of a bottleneck with a guitar fundamentally inspired Blues music. Radio experiments accidentally led to the development of the first synthesizers and electronic music, whereas current music bands using familiar objects or even vegetables to make music are renowned for their originality (e.g., Stomp, Vegetable Orchestra). Beyond the cultural-evolutionary processes by which new instruments, musical genre, and subgenre might emerge (Graham, 2006), these examples emphasize how the ability to generate new sonorities for the purpose of music seems to represent both an innate feature of human ingenuity and an essential aspect of musical creativity.

On the other hand, as societal and technological progress have created a multitude of new ways for people to engage with music (Partti & Karlsen, 2010), many promising research developments support creativity in relation to music technology (e.g., Mellor, 2007, 2008). These technological advances are viewed as stimulators of creative possibilities, allowing the realization of ideas when motor skills lag behind ideas. Taken together, it appears that these “new” dimensions of creative work in music should be captured through suitable assessment methods of creative thinking that (a) minimize the involvement of instrumental and musical mastery and (b) take into account the creative thinking engaged in sound production and computer-based composition, two facets of creative thinking in music that are closer to current musical experiences. Such a measure would extend the study of musical creativity to the general population for a better understanding of musical creativity in relation to intellectual, social, and personal development (Hallam, 2010).

The Musical Expression Test

Overview

As reviewed above, the assessment of creative thinking in music faces many challenges. Through two pilot studies and the present work, we attempted to address these issues by developing the Musical Expression Test (MET) designed for children, adolescents, and adults, with or without prior musical training (minimal technical skills involved). The MET was developed as a multimethod technique integrating the divergent-thinking measurement approach (e.g., Webster, 1994), creative process analysis in computer-based composition tasks (e.g., Folkestad et al., 1998; Hickey, 1997), and classic product-based assessment (e.g., Auh, 1997; Eisenberg & Thompson, 2003, 2011; Priest, 2006). Consistent with established models of musical creativity (Webster, 1990), and a recent framework for the domain-specific assessment of potential for creativity (Lubart, Besançon, & Barbot, 2011), the MET aims specifically to assess two sets of abilities: (a) Divergent-exploratory thinking processes—expanding the range of solutions in creative problem-solving and (b) convergent-integrative thinking processes—combining elements in new ways. These key-processes are measured through four standardized subtests (musical activities) including free exploration, alternative uses, composition, and improvisation. These subtests are chained and presented to subjects as distinct stages of an individual 1-hour-long composition workshop.

The general measurement principle of the MET involves two levels of analysis. The “product-based” level focuses on the musical piece produced in response to the composition and improvisation tasks, which involves the Consensual Assessment Technique (CAT). The “behavioral level” focuses on the exploratory behaviors enacted in response to the tasks and assesses divergent-exploratory processes through a systematic observation protocol. Rather than relying on the musical contents of the test responses (involving musical notation), this innovative approach focuses on the instrumental gestures—that is, the physical actions involved in the production of sound (Widmer & Goebi, 2004)—which are hard-wired to the musical production. This focus on physicality—inspired by Getzels’ (1964) and Getzels and Csikszentmihalyi’s (1974) methods for the study of problem-finding behaviors in visual-art creativity—allows quantitative and qualitative analyses of the exploratory activity (e.g., number, diversity, and frequency of musical events) enacted in response to the tasks, with a particular emphasis on the use of the material conditions, in an ecological perspective (cf. Gibson, 1977). This allows capturing a rarely studied facet of creative thinking in music: The ability to produce unusual sonorities based on the new or unusual (i.e., infrequent) use of objects (cf. Gilhooly, Fioratou, Anthony, & Wynn, 2007; Leman, 1999).
Materials

The MET involves a “sound-production set” including 12 elements, which are playful, easy to use, providing diverse physical properties, potential uses, and sounds. This set represents three categories, according to whether the element is mainly Melody-oriented (e.g., Glockenspiel, Kalimba), Rhythm-oriented (e.g., Tambourine, Cymbal), or “Sound-Effect” oriented, including “sound objects” (very simple elements designed to produce sound-effects or atmospheres, such as the Rainstick or Frog-guiro), and “heterogeneous objects” whose “natural” use is not to produce music or sounds (e.g., spoons, a plastic tube). Obviously, even a melody-oriented element can be used for other functions, such as for sound-effects or rhythmic sequences. Other equipment involved in the MET includes a computer-based audio-numeric sequencer software and virtual studio (Cubasis VST, Steinberg), along with an external sound-card and VST controller, a microphone, a headset, and a hi-fi system, which together allow multitrack recording and advanced orchestration possibilities. Finally, a video camcorder is used for the systematic observation procedure. The placement of the material is standardized, presented in an identical format for each subject (to control for possible original combinations between elements suggested by their proximity in placement).

The MET Procedure

The MET is divided into four complementary subtests (stages) structured as a workshop introducing composition. This structure and the MET subtests’ content (i.e., instructions, material) were selected and calibrated for duration and other parameters, on the basis of our preliminary work. Figure 1 represents the organization of the MET procedure.

As indicated, an initial presentation stage establishes the contact between the test administrator and the subject (there is no measurement involved at this point). The test administrator presents briefly the MET and asks the subject to indicate his or her familiarity with each elements of the sound-production set (so that each element is seen). After this introduction, the four subtests are administered as follows:

1. Free exploration (3 minutes). Participants are invited to explore freely the sound-production set during 3 minutes. This stage focuses on spontaneous explorations, free of restrictions, and without the presence of the test administrator who could inhibit the very first exploratory behaviors.

2. Mini-Games (10–15 minutes). The test administrator presents a musical composition recorded with the sequencer software, involving six elements of the sound-production set (recorded in six separate, layered-tracks). While familiarizing the participant with the possibilities of the sequencer further used to create his or her own composition, this stage leads to two distinct tasks: (a) identifying the element used for each recorded layer (each layer is played independently), after which the participant has to (b) propose as many alternative uses of the identified elements, to produce diverse sounds (2 minutes for each item). This alternating movement between both tasks is repeated for the six targeted elements (i.e., there are six items). There is no measurement in the first identification task, whereas the second—inspired from classic “alternative uses” tasks (e.g., Guilford, 1967)—aims to measure the ability to generate potential uses for the purpose of sound production. Resulting scores reflect individual differences in this ability, in terms of quantity (fluency), diversity (flexibility), and statistical infrequency (originality) of the generated answers with regard to those proposed in the tested population.

3. Composition (30 minutes). In this stage, the subject creates a 30-s musical piece using the sequencer presented through the introductory and the minigames stages. Specifically, the participants have the possibility to record, step-by-step, each layer (musical segments) of their composition and then to overlay the recordings in an integrated musical piece. Participants are free to use the sound-production set in any way, and to record, erase, and edit (delete, copy, and paste sections) musical sequences. However, the protocol allows a maximum of eight layers to be included in the final composition. The test administrator handles the recordings and the software to save time and facilitate the process. This subtest ends when the subject considers the composition complete, within a maximum of 30 minutes.

4. Improvisation (5 minutes). This final stage consists of recording a 1-min improvisation after a 2-min warm-up/
preparation-time before recording. This subtest is optional and is administered according to the goals (e.g., for research needs or to monitor progress in improvisation-based training) and the targeted population of the test administration. Experience shows that, even placed as the final stage, improvising without particular guidelines or constraints can be somewhat difficult for novices or subjects with low self-confidence.

**General Coding and Scoring Principles**

Beyond the CAT involved for judgments of the musical products resulting from the composition and improvisation subtests, the MET provides behavioral measures of musical explorations observed across the subtests. These measures are derived from a systematic observation protocol developed upon the transcription of 30 observations of the pilot version of the MET (using Ad libitum sampling) which led to the identification and the definition of inclusion and exclusion parameters for the coding of the exploratory behaviors. As a result, observation units are defined with regard to three criteria: (a) Behavioral orientation (element explored); (b) Behavioral structure (form of the exploratory behavior); and (c) Function (consequence of the behavior). These criteria are then combined to obtain the observation units (called “exploratory units”) referring to distinct musical explorations, which are tagged and quantifiable.

Specifically, the orientation criterion is coded using a number referring to the element of the sound-production set used (additional codes are proposed to transcribe orientations not listed in the sound-production set such as different parts of the body, table, keys, and other external sound-sources). The structure criterion is defined according to a catalogue of typical instrumental gestures identified in the pilot studies and in Mialaret’s catalogue (Mialaret, 1997), which we summarized according to gestures’ morphological similarities. Each behavior falls in one of 11 categories (e.g., hits, rubs, blows, glissandi, and sweeps) coded with one or two letters. Finally, the function is based on a very simplified but sufficiently exclusive taxonomy which includes three possible consequences of the exploratory behavior: Melody-oriented (unit including sounds with locatable pitch variation), Rhythm-oriented (unit with no pitch variations, and the presence of rhythm), and Sound-Effect oriented (emphasize sound effect, noise, ambiance, and atmosphere sounds). Additional criteria, such as operators and play-zone locators, are available to reflect complex exploratory behaviors and to set the limits of the diverse exploratory units. Using this transcription system, an exploratory unit such as “1aHM” would represent an exploratory behavior in which the subject uses the glockenspiel (orientation criterion coded “1”) by hitting (structure criterion coded “H”) with its mallets (complementary orientation criterion coded “a”), producing a series of sounds with locatable pitch (function criterion coded “M” for “Melody-oriented”).

The exploration units are then used as the basis of the computation of the MET’s behavioral scores, to quantify the exploratory activity enacted in response to the various tasks: fluency (number of exploratory units), flexibility (number of different exploratory units), and originality (relative infrequency of each exploration unit with respect to the units indexed in the experimental population for each subtest). These variables are computed for each of the MET subtests (free exploration, answers of the “alternative use” task, during the compositional activity, and during the improvisation stage). Finally, behavioral measures are also computed upon the exploration units used in the actual musical piece (composition stage) to refer to the number, the variety, and the originality of the exploration units recorded and kept in the final musical piece.²

**An Empirical Study With the MET**

The goal of this study was to examine empirically the MET and confirm its appropriateness for administration in the general population, using a sample of adolescents unfamiliar with composition. Despite the apparent face validity of the MET as a measure of targeted aspects of creative thinking in music, we sought to estimate important properties of the MET’s coding protocol and resulting test scores with a special focus on (a) evidence for reliability (including interrater reliability of the exploratory units coding system and resulting behavioral scores, internal consistency of the divergent-exploratory behavioral scores computed from the six “Alternative uses” items, and interrater agreement of the product-based assessment obtained through the CAT); (b) possible gender and musical training effects on the test scores; and (c) exploration of the convergence between the behavioral measures and the product-based assessments (including the examination of behavioral differences associated with typical “creative styles” identified and derived upon the product-based measures). The optional improvisation subtest of the MET and the compositional process data were not analyzed in this study.

**Method**

**Participants.** Forty-two adolescents (29 girls, 13 boys) in the first year of high school ($M_{age} = 15.6$ years, $SD = 0.6$) participated in the MET. Forty-five percent of the participants had prior musical experience (12 girls, 7 boys), but none of them had prior experiences of musical composition. Two independent coders (including the first author and a music teacher) were involved to transcribe the video corpus after the MET systematic observation protocol. Finally, eight experts in music (one woman and seven men), consisting of four professional composers including teachers of composition, and four amateur musicians with more than 10 years of musical experience, participated as judges for the CAT.

**Measures and procedure.** The MET was implemented in an afterschool program taking place in the high school where the participants were recruited. The program was advertised as a composition workshop for research purposes (no reference to musical creativity as a target of the study was made). Consent forms authorizing for the audio-recording and video-filming of the session were signed by the participants and their parents. Following the MET’s procedure, we administered the MET in individual sessions not exceeding one hour. Musical pieces resulting from the

² Even though we are not presenting results regarding the compositional process here, it is worth noting that the MET scoring protocol includes an extensive guideline for the coding of “Compositional units” (to transcribe the activity during the composition stage, such as recording, editing, erasing, listening) for further study on the creative process. The full coding and scoring protocol is available upon request to the corresponding author.
MET tasks (composition and improvisation) were exported and given to the participants on an audio CD.

After the test administration period, the MET video recordings were transcribed by the two coders. After a training program based on the pilot studies corpus, each coder handled independently half of the video recordings. In addition, a double-coding procedure was completed for 25% of the recordings leading to the extraction of about 1,300 observation units transcribed by both coders. This overlapping corpus was used as a basis of the interrater reliability analyses. The CAT was used to assess several criteria on the 42 musical pieces resulting from the MET composition subtest. The eight judges evaluated independently the compositions through a pilot version of the Consensual Assessment Technique-interface (CAT-i; Barbot & Pouyade, 2006–2011), an Internet-based system to facilitate the application of the CAT and its major methodological guidelines (e.g., Henninger, 1994), such as the randomization of the order of creative products for each judge, while permitting an instantaneous, remote data-collection from geographically distant judges. The compositions were assessed using a seven-point Likert scale in three independent series for three criteria defined through the interface—these definitions were based on Audh (1997), and Hickey (1999)’s rubrics: (a) Creativity (degree of both originality and coherence of the composition); (b) Technical Quality (degree to which the tonal and rhythmic elements in a composition show technical mastery in terms of tonal center and rhythmic regularity); and (c) Expressiveness (degree to which the composition is musically expressive and reflects the participant’s aesthetic sensitivity to music).

Data analyses. The interrater reliability analyses conducted on the observation corpus involved two levels. On a “qualitative” level, we compared the exploratory units coded by the raters and we computed the percentage of agreement (exact same unit coded by both raters), disagreement (variation in the coding of a unit across raters), and omissions (when one rater coded a unit while the other did not).3 On a “quantitative” level (MET’s behavioral scores resulting from the coding of the observation units), we calculated intraclass correlation coefficients (Shrout & Fleiss, 1979) between the scores computed separately from the coding of the two independent raters. The internal consistency of the fluency, flexibility, and originality scores derived from the six items of “alternative use” task (Minigames subtest) was estimated by computing Cronbach’s alpha, followed by the computation of composite scores (reflecting the performance across the 6 items). Similarly, interjudge correlations of the product-based measures and resulting Cronbach’s alpha were computed using the ratings of the eight independent judges, before deriving the consensual scores of Creativity, Technical quality, and Expressivity.

Gender and musical training effects on the test scores were then examined using multivariate analyses of variance (MANOVAs). Finally, we used a correlational approach to examine the convergence between behavioral and product-based measures, and we sought to identify “creative styles” (typical profiles observed with the product-based consensus scores) using hierarchical ascendant classification (Ward’s method) followed by MANOVAs to examine which behavioral measure would differentiate the identified groups.

Results

Reliability analyses.

Behavioral coding and scoring reliability. The percentage of agreement and interrater correlation coefficients observed for the coding of the various MET subtests are presented in Table 1. On the qualitative level, the double-coding procedure showed a satisfactory percentage of agreement, with a median of 86% of identical observation units coded by the two raters. On the quantitative level, intraclass correlation coefficients (ICC) between the score derived independently from each rater, showed a high interrater agreement (median ICC = .92, p < .001). These results are consistent across the MET subtests and variables (i.e., fluency, flexibility, and originality) and suggest the excellent reliability of the data collected in this study, as well as the general quality and suitability of the MET’s coding and scoring protocols.

Internal consistency of the alternative-uses’ scores. Internal consistency of the alternative-uses scores (that is, composite scores of fluency, flexibility, and originality, computed on the basis of the six items) was very high, with Cronbach’s alphas of .94, .94, and .96, respectively for fluency (average interitem r = .73), flexibility (average interitem r = .73), and originality (average interitem r = .81). These high reliability coefficients support the calculation of the composite scores and confirm that the six alternative-uses items involve the same ability.

Product-based assessment reliability. Interjudge agreement for the three product-based criteria estimated with Cronbach’s alphas indicated an acceptable consensus for Creativity (.70), excellent consensus for Technical quality (.89), and limited agreement for Expressivity (.64). After computing the composite product-based scores by averaging the individual judge’s ratings for each criterion, we examined their intercorrelation. Creativity was relatively independent from technical quality (r = .38, p < .05) but strongly related to expressivity (r = .77, p < .001). Expressivity was also moderately related to technical quality (r = .58, p < .001).

Convergence between behavioral and product-based measures.

Correlational approach. Table 2 presents the intercorrelations between the behavioral scores obtained at each MET subtest and the experts’ evaluations of the musical piece resulting from the composition subtest. Moderate correlations (ranging between r = .34 and .55) were observed between the behavioral measures of the free exploration subtest and the Creativity and Expressivity of the

---

3 We used an improved version of the Miles and Huberman (1994) formula to treat errors attributable to coding omissions, to calculate the percentages of agreement (PA) between the two coders. This formula is: PA = NA/(NA + ND + (NO1 + NO2/2))×100 (where PA = percentage of agreements; NA = number of agreements; ND = number of dissensions; NO1 = number of omissions for observer 1; NO2 = number of omissions for observer 2). We did not correct PA for chance using Cohen’s Kappa (based on the probability of occurrence of the observation units) because we assume that PA reflects only a minimal, even null influence of chance, as there exists conceptually an infinite number of exploratory units.

4 Whereas behavioral variables were approximately normally distributed, we used Spearman rank-order correlation because of the small number of participants in this study. Nevertheless, analyses using Bravais–Pearson correlation coefficients yielded a highly similar pattern of results.
composition (product-based scores). The Technical quality of the composition was not related to the behavioral scores obtained in this subtest.

Interestingly, the behavioral measures obtained in the subsequent subtests appeared to be rather independent from the product-based measures. Only the number of proposed alternative uses (fluency) in the minigame subtest was weakly correlated with the technical quality ($r = .38, p < .05$) and expressivity ($r = .37, p < .05$). Finally, the results indicated a moderate convergence between the behavioral measures extracted from the musical piece corpus and the product-based measures. Fluency, flexibility, and originality of the explorations units incorporated in the musical piece were positively correlated with the Creativity of the musical piece (with correlation coefficients ranging from $r_1 = .34$ to $r_2 = .41$), whereas the same variables were negatively correlated with the Technical quality of the musical piece ($r_3 = -.39$ to $r_4 = -.48$), as assessed by the experts.

**Training and gender effects.** A set of analyses was conducted to investigate possible effects of prior musical training and gender on the behavioral and product-based measures. Gender was not associated with individual differences on the three product-based scores, $F(3, 36) = 2.6, p = .18$, whereas a slight effect of the prior musical training was obtained, $F(3, 36) = 3.34, p = .02, \eta = .22$. Specifically, Tukey’s HSD post hoc tests indicated that participants with prior musical training scored significantly higher than participants with no prior training on the Technical quality measure ($p < .05$). No significant differences on Creativity and Expressivity measures were observed, and there were no interactions between gender and musical training. Analyses performed on the behavioral measures did not indicate main effects of gender or prior musical training except for the flexibility scores in the free exploration and composition subtests in which participants with prior musical training tended to score lower than participants with no prior training ($p < .05$). No other effects on the behavioral measures were observed.

**“Creative style” of the musical pieces and their behavioral correlates.** The Hierarchical Ascendant Classification was used to identify groups of individuals with similar profiles (interpreted as “creative styles”) based on their musical piece’s scores of Creativity, Technical quality, and Expressivity. This analysis allowed the identification of four homogeneous groups of individuals, summarizing common profiles in our data with high parsimony ($F[9, 87.8] = 22.9, p < .001, \eta = .70$). Table 3 presents the descriptive statistics of the four groups (with results of the univariate analyses), and Figure 2 displays the average profiles of these groups.

A MANOVA followed by Tukey’s HSD post hoc tests indicated the extent to which the four groups were significantly different on the three product-based scores. Based on the most salient aspects of the four identified groups, we labeled them as follows: The gifted group (20% of the sample) showed significantly higher scores than the others on the three product-based scores; the creative playful group (41%) corresponds to compositions considered to be fairly creative and expressive but low on Technical quality; the academic group (24%) reflected compositions with average levels of Expressivity and Technical quality but weak creativity scores; finally, the disorganized group (15%) was associated with low Creativity, Technical quality, and Expressivity scores, suggesting a lack of coherence in the musical contents. The MANOVA did not indicate any interaction effects between these creative-styles and the other group-factors taken into account in this study (i.e., gender and prior musical training).

Finally, a MANOVA on the behavioral measures using the creative-style group as an independent variable revealed a significant effect, $F(3, 81) = 1.68, p = .03, \eta = .43$. Post hoc tests indicated that the gifted group (associated with the highest product-based scores) differed from the disorganized group (regarded as less creative) by a greater quantity ($p = .03$) and variety ($p = .02$) of instrumental explorations (i.e., exploratory units) during the free exploration stage. The creative playful group (moderate Creativity associated with limited Technical quality) compared with the academic group (average Technical quality and limited Creativity) showed higher scores of fluency ($p = .02$) and originality ($p < .01$) in the behavioral measures extracted from the musical piece. In other words, the participants in the creative playful group tended to use more exploratory events and to generate less conventional uses of the sound-production set in their

### Table 1

<table>
<thead>
<tr>
<th>Qualitative level</th>
<th>Percentage of agreement</th>
<th>Free exploration</th>
<th>Stage Games</th>
<th>Composition</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluency</td>
<td>.93</td>
<td>.87</td>
<td>.98</td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td>Flexibility</td>
<td>.94</td>
<td>.95</td>
<td>.98</td>
<td>.96</td>
<td></td>
</tr>
<tr>
<td>Originality</td>
<td>.94</td>
<td>.89</td>
<td>.91</td>
<td>.91</td>
<td></td>
</tr>
</tbody>
</table>

* Intra-class correlation coefficients significant at $p < .001$.

### Table 2

<table>
<thead>
<tr>
<th>Behavioral measures</th>
<th>Creativity</th>
<th>Technical quality</th>
<th>Expressivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free exploration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>.51***</td>
<td>.13</td>
<td>.54***</td>
</tr>
<tr>
<td>Flexibility</td>
<td>.55***</td>
<td>.11</td>
<td>.52***</td>
</tr>
<tr>
<td>Originality</td>
<td>.34*</td>
<td>.10</td>
<td>.38*</td>
</tr>
<tr>
<td>Games (alternative uses)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>.22</td>
<td>.38*</td>
<td>.37*</td>
</tr>
<tr>
<td>Flexibility</td>
<td>.27</td>
<td>.28</td>
<td>.30</td>
</tr>
<tr>
<td>Originality</td>
<td>.30</td>
<td>.21</td>
<td>.28</td>
</tr>
<tr>
<td>Composition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>.00</td>
<td>.30</td>
<td>.12</td>
</tr>
<tr>
<td>Flexibility</td>
<td>-.03</td>
<td>.17</td>
<td>.04</td>
</tr>
<tr>
<td>Originality</td>
<td>.00</td>
<td>.14</td>
<td>.09</td>
</tr>
<tr>
<td>Musical Piece</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>.34*</td>
<td>-.47**</td>
<td>.08</td>
</tr>
<tr>
<td>Flexibility</td>
<td>.41**</td>
<td>-.48**</td>
<td>.09</td>
</tr>
<tr>
<td>Originality</td>
<td>.40**</td>
<td>-.39*</td>
<td>.19</td>
</tr>
</tbody>
</table>

* $p < .05$. ** $p < .01$. *** $p < .001$. 

---

CREATIVE THINKING IN MUSIC
musical piece. Behavioral measures obtained during the minigame subtest and the composition subtest did not significantly differentiate the four creative styles.

Discussion

This first empirical evaluation of the MET had several objectives. First, we sought evidence of various aspects of the MET’s reliability. Second, we wanted to examine the effects of prior music training on the MET scores and gauge MET’s suitability for applications in the general population. Finally, our third goal was to explore the relationships between the MET’s behavioral indicators and product-based assessment through a correlational approach and a complementary examination of “creative styles” and their behavioral correlates. Our results are discussed separately in line with these objectives, and some limitations of this study and directions for future work are presented.

Evidence for reliability. The results showed high interrater agreement between two independent coders and the resulting scores derived from the observation corpus. This suggests that the coding and scoring protocols used in the MET are suitable and lead to robust and reliable scores. Similarly, we observed high internal consistency of the alternative-uses scores, confirming that the items in this subtest measure the same ability to propose numerous, various, and original uses of the sound material. Finally, the CAT used for the assessment of several qualities of the musical piece, resulting from the composition subtest, yielded acceptable inter-judge agreement between eight experts. As a whole, this evidence highlights the strengths of the MET coding and scoring principles and the overall quality of the data obtained in the study.

As the levels of expertise of the judges might interfere with creativity ratings and previous research has shown that composers tend to score musical composition in an idiosyncratic way as compared with other groups of experts (Hickey, 2001), further analyses regarding product-based assessment are needed. This focus is important as the product-based scores are used as the principal criterion to evaluate the behavioral indicators and as the basis of the four creative styles in our data. Would similar product-based scores be generated by different groups of experts (e.g., music teachers)? To what extent did the levels of expertise of the judges involved in this study influence their judgment of the

Table 3
Descriptive Statistic of the Four Creative Styles on the Product-based Measures and Result of the Univariate Analyses

<table>
<thead>
<tr>
<th>Creative style</th>
<th>Gifted M (SD)</th>
<th>Creative playful M (SD)</th>
<th>Academic M (SD)</th>
<th>Disorganized M (SD)</th>
<th>F(3, 39)</th>
<th>p</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creativity</td>
<td>1.24 (.55)</td>
<td>.22 (.58)</td>
<td>−.73 (.78)</td>
<td>−1.13 (.10)</td>
<td>26.5</td>
<td>&lt;.001</td>
<td>.67</td>
</tr>
<tr>
<td>Technical quality</td>
<td>1.22 (.82)</td>
<td>−.76 (.48)</td>
<td>.57 (.52)</td>
<td>−.70 (.30)</td>
<td>31.3</td>
<td>&lt;.001</td>
<td>.71</td>
</tr>
<tr>
<td>Expressivity</td>
<td>1.60 (.49)</td>
<td>−.26 (.52)</td>
<td>−.30 (.53)</td>
<td>−1.13 (.21)</td>
<td>46.1</td>
<td>&lt;.001</td>
<td>.78</td>
</tr>
</tbody>
</table>

Figure 2. Average creative style profiles.
musical compositions? Future work focusing on the judgment of
the creative products resulting from the MET could help to answer
these questions and to better target (a) the appropriate group of
experts to judge the creativity of the musical compositions, as well
as (b) the optimal criteria to be used for the assessment of the
musical compositions (e.g., melodic originality or rhythmic coher-
ence).

Musical training and technical mastery effects. Results
indicated that product-based Creativity scores were somewhat
independent of Technical quality scores, which corroborates the
results of several studies (e.g., Auh, 1997; Webster, 1987) under-
lining how “general aptitudes” in music should not be confounded
with the ability to think creatively in music (e.g., Webster, 1990,
1992). Accordingly, prior musical training was associated with
relatively higher scores of Technical quality on the musical piece
but no effects on the Creativity scores. This confirms that “musi-
cally correct” compositions (i.e., technically mastered) are not
necessarily more creative (e.g., Auh & Walker, 1999), as also
illustrated by the academic and the creative playful styles.

On the behavioral level, participants without prior musical train-
ing did not show more explorations (fluency) than those with prior
training contrary to the Seddon and O’Neill’s (2003) study. How-
ever, they showed more flexible uses of the sound-production set
than the latter during the compositional process. This tendency
to focus on limited alternative use of the sound-production elements
may reflect a preference for “in-depth” musical explorations and/or
the repetition of musical segments among the participants
with musical training, rather than a focus on “in-breadth” musical
explorations (associated with a greater diversity of instrumental
explorations). Indeed, relatively lower flexibility (as a behavioral
indicator of the diversity of the exploratory activity) might indicate
a higher focus on a given element of the sound-production set,
which may ultimately result in a better technical “mastery” of the
given element. This is supported by the negative correlations
between the behavioral measures extracted from the musical com-
position and the Technical quality scores. This result helps clarify
why the higher Technical quality scores of the group with prior
musical training are also associated with lower flexibility during the
composition task. As a whole, these results stress the need for
further analyses using process data with a specific focus on “rep-
etition” (e.g., Kratus, 1989). This could contribute to understanding
some of the overlaps between specific behavioral- and product-
based measures in the MET, which may be explained through the
mediating effect of prior musical training. However, the patterns of
results obtained in this study, along with anecdotal observations
during the administration phase of the MET, confirmed its suit-
able use for the general population. This supplements recent
efforts to study musical creativity in this population using increas-
ingly popular computer-based compositional approaches (e.g.,
Mellor, 2007).

Links between MET’s behavioral indicators and product-
based assessments. Consistent with results described in the
problem-finding literature in music (e.g., Brinkman, 1999), the
quantity, variety, and originality of the explorations in the free
exploration subtest were related to the creativity of the musical
piece produced in a later stage of the MET (that is, the composition
subtest). Indeed, the convergence observed between the free ex-
ploration scores and the product-based measures (up to 30% of
shared variance) suggests that the free exploration subtest could
correspond to a “problem-finding” period for the composition
stage, in which participants would “draft” their musical piece. This
hypothesis is consistent with the notion that a creative musical
piece must have intention or a plan (e.g., Hickey & Webster,
2001). Alternatively, it is possible that the quantity of the explor-
atory behaviors observed during this free exploration stage reflects
intrinsic motivation toward exploring the material conditions,
which contributes to musical creativity. Similarly, it is possible
that score differences among participants in this subtest may
reflect how familiar and comfortable they are with the various
aspects of the MET settings, which could ultimately intervene in
the compositional process and the creativity of the resulting mu-
sical piece. Future research to address this possibility is needed,
examining further the extent to which motivation and familiarity
interacts with the MET and is empirically related to problem-
finding behaviors in creative problem-solving in music.

Interestingly, the behavioral scores obtained through the com-
position subtest and the alternative-uses task appeared relatively
independent of the product-based creativity score. These results
indicate the need to examine the links between the composition
subtest scores and the product-based scores in light of “process
data” (e.g., is a significant amount of time spent on developing a
musical idea, or repeating a musical segment?). The results also
suggest that the alternative-uses scores reflect presumably a facet
of musical creativity that is not captured in the product-based
measures, or which is not involved in compositional creativity. To
this regard, it is relevant to include this alternative-uses task in the
MET so that it covers a wider sample of creative behaviors in
music. However, future examination of this task is also needed to
establish its external validity with a relevant criterion such as a
classic “alternative-uses” task in the verbal domain (e.g., Guilford,
1967). This would allow us to gauge whether this facet of musical
creativity represents a more general divergent-exploratory ability
that consists of generating many, varied, and original uses of
objects (Gilhooly et al., 2007), also involved in other domains of
creative work.

Finally, the convergence between product-based measures and
behavioral scores extracted from the actual musical composition
provided some evidence of the importance of musical exploratory
behaviors for creativity. Specifically, our results suggested that
“musical completeness” could be an important factor for composi-
tional creativity as a high number of exploratory events in the
musical content was associated with higher creativity scores. Fur-
thermore, the use of unusual sounds (reflected by rare exploratory
units, that is, high originality scores) was related to the creativity
of the musical piece, confirming the importance of this aspect of
creative thinking in music (cf. Leman, 1999). Conversely, it ap-
ppeared that these behavioral indicators—number and originality
of the exploratory events—were negatively related to the Technical
quality of the musical piece. It seems consistent that the integration
of a large variety of sound elements in the musical piece will
increase the difficulty to achieve coherence between these ele-
ments and to ensure technical control. As suggested by the gifted
style (reflecting high levels in all product-based scores), it appears
that both musical completeness and originality of the sound ele-
ments used in the musical piece must be associated with a general
technical mastery to result in a highly creative product.

Study limitations and future directions. The study pre-
sented here provided the first evidence for the robustness of the
MET to measure various aspects of creative thinking in music through a sound and reliable coding and scoring system. This standardized, multimethod framework proved to be applicable to both musically trained and untrained participants, which suggests great potential for future research on musical creativity. However, some limitations of this study have been pointed out, and some others must be noted here and addressed in future research on musical creativity with the MET.

A first point concerns the limited sample size and the focus on specific goals in this study: we must recognize that the data obtained are not sufficient to examine completely the psychometric properties of the MET, although our first results are promising. Future work involving larger samples is needed to establish other psychometric properties such as test–retest reliability, external and internal validity, as well as for the development of proper norms for the application of the MET in research/educational settings. The focus on external validity is crucial, in particular for the product-based assessment of the musical piece resulting from the composition subtest. Indeed, by simulating creative work in music composition with specific task demands and constraints (e.g., demands placed on the MET’s test administrator, material conditions, time limits), we may be restricting the creative ability measured to this specific task. This is in fact a common problem of any product-based/integrative task of creative thinking. According to confluence models, a person’s profile across several components involved in creative work may fit more or less closely the requirements of a given task, and this leads to variations in the level of creative performance across domains and also across tasks within domains (Lubart & Guignard, 2004). Future research must explore the extent to which the specific requirement of the subtests in the MET capture the key features involved in the broader ability to think creatively in music (if such a general ability exists). This would help to determine whether some features of the MET tasks are undesirable or should be adjusted.

In another line, the identification of the four “creative styles” in our data provided the first evidence that different behavioral approaches with the MET’s sound-production set were associated with different qualities in the musical content (e.g., the creative playful, vs. the academic styles). This typology appears useful to describe, albeit broadly, various approaches in musical composition, but more replication work is needed to establish typological stability (can we identify the same “styles” with other samples?) and to set the parameters of a robust classification in these groups. Finally, we must acknowledge that the sophistication of the coding and scoring system in the MET may reduce its applicability. Present work is ongoing, in view of a simplification of the coding and scoring system, without substantial loss of information.

**Conclusion**

To date, the majority of the research on creative thinking in music has targeted well-adjusted children with musical training, limiting the theoretical advances and possible practical implications of this relevant topic to music education concerns. However, as the positive effect of the musical activity is pointed out beyond its role in the music-classroom (e.g., Hallam, 2010), musical creativity appears to be a relevant trigger of change in up-to-date therapeutic and educational programs, in phase with the current musical experiences of today’s youth. Consequently, treating musical creativity as a facet of human creative potential that can be nurtured and studied in the “general population” represents an advance for creativity research and its applications. Although such a premise is impeded by many challenges related to the musical domain, researchers must “compose” with these issues and pursue efforts to improve our knowledge on this topic, which requires better assessment methods for musical creativity, as a basic condition.

Toward this effort, we presented the MET, a new multimethod assessment of creative thinking in music with a focus on instrumental exploratory behaviors. Although the sample size in this study was restricted, encouraging results regarding reliability and convergence between the behavioral measures and classic product-based assessment were obtained. Built upon increasingly popular computer-based compositional approaches (e.g., Mellor, 2007), the MET proved to be suitable to be administered to (and by) individuals without any prior musical training. Future research efforts are needed to explore in depth the MET’s psychometric properties using larger samples, and including analyses of the criterion validity—even though the lack of any validated measure of musical creativity might limit this investigation. Beyond the MET’s apparent soundness and face validity, such research efforts are essential to determine whether the MET constitutes, at this point, a serious basis for future research on musical creativity.

**References**


Brinkman, D. (1999). Problem finding, creativity style and the musical...


Ward, C. J. (2009). Musical exploration using ICT in the middle and...


Received October 6, 2011
Revision received January 10, 2012
Accepted January 10, 2012