http://www.diva-portal.org

This is the published version of a paper published in *Personality and Individual Differences.*

Citation for the original published paper (version of record):

Ullén, F., Mosing, M., Holm, L., Eriksson, H., Madison, G. (2014)
Psychometric properties and heritability of a new online testfor musicality, the Swedish Musical Discrimination Test.
Personality and Individual Differences, 63: 87-93

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-86650

Personality and Individual Differences 63 (2014) 87–93

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Psychometric properties and heritability of a new online test for musicality, the Swedish Musical Discrimination Test

Fredrik Ullén [a,*], Miriam A. Mosing [a], Linus Holm [b], Helene Eriksson [a], Guy Madison [b]

[a] *Dept. of Neuroscience, Karolinska Institutet, SE-171 77 Stockholm, Sweden*
[b] *Dept. of Psychology, Umeå University, SE-901 87 Umeå, Sweden*

A R T I C L E   I N F O

Article history:
Received 21 November 2013
Received in revised form 22 January 2014
Accepted 25 January 2014

Keywords:
Music
Sensory discrimination
Expertise
Psychological testing
Training

A B S T R A C T

We examine, in 6881 twin individuals, the psychometric properties of a new test (the Swedish Musical Discrimination Test, SMDT) that was developed to tap auditory discrimination of musical stimuli. The SMDT consists of three subtests measuring discrimination of melodies, rhythms, and single pitches, respectively. Mean test taking times for the subtests were 3.0–4.6 min. Reliability and internal consistency were good with Cronbach's alpha values and Spearman–Brown split-half reliabilities between .79 and .89. Subtests correlated positively (r values .27–.41). Criterion validity was demonstrated in three ways: individuals that had played a musical instrument scored higher than individuals that had not (Cohen's d .38–.63); individuals that had taken music lessons scored higher than individuals that had not (Cohen's d .35–.60); finally, total hours of musical training and SMDT scores correlated (r values .14–.28) among those participants that had played an instrument. Lastly, twin modelling revealed moderate heritability estimates for the three sub-scales. We conclude that the SMDT has good psychometric characteristics, short test taking time, and may serve as a useful complement to existing tests of musical ability.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

1. Introduction

Music is a human universal of profound significance for most people. At the same time, musicality, broadly defined as the capacity to learn and perform music-related tasks, seems to vary substantially between individuals. There have been many endeavours to objectively measure musicality since the early 20th century. Several standardised, explicit forms of musicality tests have been constructed, both with practical aims, such as selection of students for musical training, and for research purposes (Boyle & Radocy, 1987; Shuter-Dyson, 1999; Shuter-Dyson & Gabriel, 1981; Wallentin, Nielsen, Friis-Olivarius, Vuust, & Vuust, 2010). Exactly how musicality is measured can make substantive differences to the information one obtains and hence to the aspects of the phenomenon one can study. Correlations between different tests and between tests and criteria such as teacher's ratings and music school grades tend to be in the range of .4–.6 (Shuter-Dyson & Gabriel, 1981). One important reason for these relatively moderate correlations is that different tests use different operationalisations of musicality. Indeed, within musicality testing there are two strong traditions, which differ in various characteristics. These are the 'atomistic' tradition of Seashore and the 'omnibus' approach of Wing (Jacobs, 1960; Seashore, 1919, 1938, 1947; Shuter-Dyson & Gabriel, 1981).

The atomistic approach is based on the assumption that musicality is made up of several relatively narrow and distinct musical abilities. This leads to an expectation of statistical independence (Gordon, 1969; Seashore, 1919) or at least low intercorrelations (Seashore, 1947) between tasks that tap into different abilities. Tests in this tradition have typically focused on basic sensory abilities, such as discrimination of various musically relevant sound stimuli. Empirical data consistently show moderate positive intercorrelations between discrimination tests (Carroll, 1993). While this to some degree supports the idea of independence of musically relevant perceptual abilities, individual differences in discrimination tasks are thus also influenced by more general factors. In fact, auditory discrimination tasks positively correlate with a broad range of non-musical cognitive tasks and psychometric modelling shows that general intelligence (g) is an important factor underlying the positive covariation between different 'atomistic' tests of musical discrimination (Helmbold, Troche, & Rammsayer, 2007; Lynn & Gault, 1986; Spearman, 1904; Troche & Rammsayer, 2009).

In contrast, in the omnibus approach to musicality testing, musicality is considered a general high-level ability. Tests

---

[*] Corresponding author. Address: Dept. of Neuroscience, Retzius v 8, Karolinska Institutet, SE-171 77 Stockholm, Sweden. Tel.: +46 8 52483268.
E-mail address: Fredrik.Ullen@ki.se (F. Ullén).

http://dx.doi.org/10.1016/j.paid.2014.01.057
0191-8869/© 2014 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

developed within this tradition are less concerned with characterising components of musicality but rather tend to use a holistic approach where complex, acquired musical knowledge is assessed. Typical test items may involve quality judgments of musical performances or the production of musically meaningful responses to stimuli (for example the tests of Wing and Révész; see (Jacobs, 1960)). These general differences between the two traditions also mean that the omnibus tests typically are aimed at practicing musicians, while the atomist tests can be used for a wider range of purposes. It should be pointed out that there are musicality tests that do not easily fit into either of these main traditions as they focus on musical engagement, motivation and interests rather than the cognitive capacity to process musical information as such, e.g. the Music Use Questionnaire (Chin & Rickard, 2012).

Here, a new test of musicality (Swedish Musical Discrimination Test, SMDT) is presented and its psychometric properties are analysed. The purpose of this new test is to provide measures of basic aspects of musical ability operationalised as discrimination ability for auditory musical stimuli, and the test thus continues the 'atomistic' tradition of Seashore sketched above. The SMDT consists of three subtests, Melody, Rhythm, and Pitch, which measure discrimination of rhythms, melodies, and single pitches, respectively. The three subtests are somewhat similar to the Tunes, Rhythm, and Pitch tests of Bentley (Lynn, Wilson, & Gault, 1989), the corresponding subscales of the Musical Ear Test (Wallentin et al., 2010), and the Profile of Musical Perception Skills (Law & Zentner, 2012). We aimed to design an instrument that has short test-taking time, allows for online administration, and has a suitable difficulty level for general musically untrained populations in industrialized countries. The present paper reports on the basic psychometric properties of the SMDT, including selection bias, correlations between subtests, and reliability, as well as genetic and environmental influences on each of the sub-tests based on data derived from an online administration to a larger cohort of twins. To account for the relatedness of the twins we used a randomized two-sample design, where the original sample was randomly split into two independent subsamples, in such a way that twins in the same pair were always allocated to different subsamples.

2. Methods

2.1. Participants

The participants were twins recruited from the Swedish Twin Registry (Magnusson et al., 2013). Zygosity was determined based on questions about intra-pair similarities. These have subsequently been confirmed in 27% of the twins in the registry using genotyping, showing that the questionnaire based zygosity determination was correct for more than 98% of twin pairs (Lichtenstein et al., 2002, 2006). They took the SMDT as part of a larger survey that was administered online and included numerous other questionnaires, e.g., on musical experience, personality, motivation, and interests. In total, 32,005 individuals were invited to participate, and 11,543 logged in on the questionnaire website. The present analyses are based on data from 6881 participants, i.e., the 6718 participants that completed all three SMDT subtests, as well as another 163 participants that completed only one or two of the subtests. The sample contained 1362 full twin pairs. Of these, 711 were monozygotic (MZ; identical) and 651 dizygotic (DZ; non-identical). The participants were aged between 27 and 54 (mean = 40.7, SD = 7.7); 57.6% of the participants were female. The relatively high number of drop-outs reflects two factors: (1) the SMDT was administered close to the end of the online test battery which took between 50 and 120 min to complete; (2) the SMDT required multi-media software to be installed and function on the respondent's computer, and this was not possible for some participants.

Pilot testing of longer versions of the Melody and Rhythm subtests, for item selection purposes, were performed on a smaller sample ($n = 49$; 36 females), mainly consisting of students (age 27.8 ± 9.1 years; mean ± SD) recruited through the website Studentkaninen (www.studentkaninen.se) – a Swedish website for research volunteers.

2.2. Materials

The SMDT is composed of three subtests: Melody (18 items), Rhythm (18 items), and Pitch (27 items). In all test items, the task of the participant was to discriminate between two consecutively presented stimuli. Each subtest is constructed so that items become progressively more difficult. Total test taking times for the three subtests were 4.6 ± 1.1 (Melody), 3.2 ± 1.2 (Rhythm), and 3.0 ± 1.3 minutes (Pitch).

2.2.1. Melody

Stimuli in this subtest consisted of isochronous sequences of piano tones. The piano tones were taken from the Kontakt sound library (Steinberg AG). The pitches of the tones ranged from C4 to A#5 (American Standard Pitch; 262–932 Hz). The time interval between tones in a stimulus sequence was always 650 ms. The number of tones per stimulus increased from four to nine as the subtest progressed. For each of these six stimulus lengths, there were three items. Detailed information on the construction of the tone sequences and the selection of the final set of items is provided in the next section. The two stimuli of each item were separated by 1.3 s of silence. The pitch of one randomly selected tone was always different in the second stimulus as compared to the first stimulus. Examples of stimulus pairs for items with a stimulus length of four and nine tones are given in Fig. 1A and B. The sequence was graphically depicted as a straight horizontal line of dots which changed colour when the corresponding tone was played. The task of the participant was to indicate which tone in the second melody was different from the first. Responses were given either by pressing the computer key corresponding to the ordinal number of the differing note, or by clicking on the corresponding dot with the mouse pointer.

2.2.2. Rhythm

In Rhythm, stimuli consisted of rhythmic sequences of brief sine tones. The sine tones were 500 Hz sine waves with a total duration of 60 ms. The loudness of the tone was constant during the first 30 ms and then decreased linearly to 0 db. The inter-onset intervals between tones within a stimulus sequence had a duration of 150, 300, 450, or 600 ms. The number of sounds in each stimulus increased from five to seven as the subtest progressed, with six items for each number of sounds. The two stimuli of each item were separated by 1 s of silence. In 11 out of the 18 Rhythm item the two stimuli differed. In the remaining seven Rhythm items the two stimuli were identical. Further details on the construction of the stimulus sequences and the selection of the final set of items are provided in the next section. Examples of stimulus pairs for items using different stimuli with a sequence length of five and seven notes, respectively, are shown in Fig. 1C and D. The task of the participant was to judge whether the two stimuli were the same or different. Responses were given by pressing either one of two keys on the keyboard or by clicking one of two icons with the mouse pointer.

2.2.3. Pitch

In the Pitch sub-test, stimuli consisted of sine tones. Each tone had total duration of 590 ms and started with a 30 ms ramp from

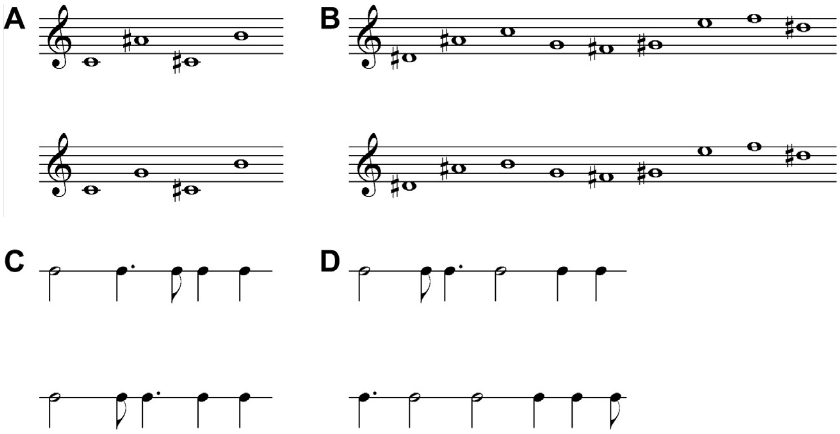

Fig. 1. Stimuli used in the Melody and Rhythm subtests. (A and B) Example of stimulus for items with four (A) and nine (B) notes for the Melody subtest. The first stimulus is shown above the second stimulus. The pitch of one note is altered in the second stimulus as compared to the first stimulus. (C and D) Example of stimulus pairs for items with four (C) and six (D) tones for the Rhythm subtest.

0 dB followed by 530 ms steady state and ending with a 30 ms ramp down to 0 dB. The frequency of one of the stimulus tones in an item was always 500 Hz. The frequency of the other tone varied in the range 501–517 Hz. The order of the two tones varied randomly so that the standard tone (500 Hz) was presented first in some items and last in the others. The two tones in an item were separated by 1 s of silence. The task of the participant was to indicate whether the first tone was higher or lower in pitch than the second tone. Responses were given either by pressing the 'H' or 'L' key on keyboard, or by clicking on corresponding icons with the mouse pointer. Difficulty increased progressively by making the pitch difference between the notes smaller in eight steps, in the following order: 17, 12, and 8 Hz difference (three items of each); 5, 4, 3, and 2 Hz difference (four items of each); 1 Hz difference (two items).

2.3. Item construction and selection

The 18 items for Rhythm and Melody were selected from a larger item pool of 36 rhythmical items and 36 melodical items based on data from a pilot test in a smaller student sample ($n = 49$).

Stimuli for the Melody subtest consisted of tone sequences with four to nine elements. The original set contained six items of each length. The first stimulus of each item consisted of a sequence of tones in the range C4–B5 that was generated randomly with the following constraints: (i) each pitch occurred only once; (ii) not all pitches in the sequence belonged to the same tonal (i.e., major, ascending minor, descending minor, or harmonic minor) scale, and (iii) all intervals between consecutive tones were smaller than one octave. The second stimulus was identical to the first, except for the pitch of one randomly selected note which was changed randomly. The same constraints (i and iii) applied to the second stimulus. An additional fourth constraint was that the random alteration of a single note in the second stimulus should not change the melodic contour of the original sequence. This procedure ensured that none of the stimuli in Melody had a clear tonality, while at the same time large leaps (i.e., an octave or more) that may induce a sense of break in the melodic line were avoided (Fig. 1A and B). The main reason for controlling that all stimuli were atonal was that a random mixing of tonal and atonal sequences could introduce large and uncontrolled variations in item difficulty: tonal melodies can be considerably easier to encode than atonal melodies of the same length (Schulze, Dowling, & Tillmann, 2012).

Rhythm stimuli consisted of rhythmic sine tone sequences with five to seven elements, containing 12 items of each length. Four different stimulus inter-onset intervals were used in the sequences: 150, 300, 450, and 600 ms. All durations were thus low integer multiples of 150 ms. These durations were combined in different ways to ensure that there was a variation in metrical structure among the sequences with a given sequence length. Stimulus pairs within the same item typically differed in that one note was moved in time in the second stimulus as compared to the first stimulus, and/or that a different starting point in the sequence was used in the second stimulus. Examples of rhythmic stimuli are given in Fig 1C and D.

The rationale of the item selection procedure for Melody and Rhythm was to identify a subset of 18 items providing as similar results as possible to the full set of items. First, scores, i.e. number of correct responses, were calculated from all possible subsets of 18 items. Secondly, these scores were compared to the total score based on all 36 items using the sum of squared differences in score as a metric. The subset which had the lowest sum of squared differences to the original set of 36 items was selected for the final test. It can be noted that the final item set for Melody did not include the pitch B5; hence, the final range of employed pitches in this test is C4-A#5. Correlations between final 18-item scores and 36-item scores were $r = .97$ for Melody and $r = .90$ for Rhythm.

Items for the Pitch sub-test consisted of pairs of tones that differed in frequency. The final set of 27 items was chosen from a larger pool of 52 items, with a range of frequency differences between 1 and 17 Hz, to achieve a suitable difficulty level and discrimination in the pilot sample (5–27 items correct, mean = 18.4).

2.4. Testing procedure

Participants received an invitation by letter containing a web address and a password, and used their own computers to access the online test. The time limit to respond was 30 days; upon request this was extended for another few days for some respondents. Data collection was undertaken during six months 2012–2013.

The complete survey featured several different psychological measures. Responses were saved continuously and the respondents could proceed at their own pace and at different times until they had finished. However, some specific tests, such as the SMDT subtests had to be completed in one session once initiated by the participant. For each subtest, an instruction was displayed providing information about the task, how responses should be entered, and how much time was allowed. This was followed by an example. All respondents did the subtests in the same order.

The aims of the project were explained to the participants in the invitation letter. Informed consent was given by each participant before data gathering begun. The study was approved by the Regional Ethical Review Board in Stockholm (Dnr 2011/570-31/5, 2012/1107/32).

2.5. Musical training and education

Data on musical training and education were extracted, for validation purposes, from a larger questionnaire on musical engagement. The following variables were used:

2.5.1. Musical activity

Participants were asked whether they play or have played a musical instrument or sung actively. This was used as a categorical variable to separate musically active and non-active participants.

2.5.2. Musical education

Participants were asked about whether they had received music lessons or not, in four different age periods (age 0–5, 6–11, 12–17, and 18-now, all in years). From this data, a categorical variable was created to separate those participants who never took music lessons from those who took lessons during some period of their life.

2.5.3. Musical training

Musically active participants were asked subsequent questions about their practicing intensity, i.e., average hours of training per week during different age periods (the same age periods as for musical education), as well as their starting and ending (where applicable) year of musical training. From this raw data an estimate of the total life-time hours of musical training was calculated for each musically active participant.

2.6. Statistical analyses

All results reported here are based on the twin data. To account for the relatedness of the twins in the non-genetic analyses, we used a randomized two-sample design, splitting the sample by assigning the members of complete pairs randomly to two independent subsamples: Sample 1 ($n = 3465$) and Sample 2 ($n = 3416$). All analyses were performed separately in both samples.

Means, standard deviations and ranges of scores are presented for the three subtests of the SMDT. Homogeneities of the subtests were estimated by computing Cronbach's alphas and Spearman–Brown split-half coefficients (odd versus even items). Given the dichotomous nature of the data, these analyses were based on tetrachoric correlations. For validation purposes, SMDT scores were compared between musically active and non-active participants, as well as between musically educated and non-educated participants, using two sample t tests. For the musically active participants, Pearson correlations between SMDT scores and estimates of total hours of musical practicing were calculated.

2.7. Twin modelling (using the full sample including single twins)

The classical twin design makes use of the differences in genetic sharing between MZ and DZ twins. While the former share 100% of their segregating genes, the latter share only 50% on average. This information can be used to partition trait variance into that due to *additive genetic* (A) and environmental influences (*common* (C) – all influences shared between the pair making them more alike and *unique* (E) – all influences not shared between the twins making them more different from each other including measurement error). With the help of structural equation modelling the combination of ACE influences best explaining the population variance in a trait can be estimated. Here, maximum likelihood (ML) modelling procedures using the flexible matrix algebra program Mx (Neale, Boker, Xie, & Maes, 2006) were used to derive parameter estimates for the saturated model. We explored age and sex effects on the mean and for normality of the three musical sub-domains prior to genetic modeling.

3. Results

3.1. Reliability and validity of the SMDT

All results are presented separately for the two independent subsamples. Descriptive statistics on test scores, as well as data on the internal consistency (Cronbach's alpha) and the reliability (Spearman–Brown split-half coefficient) of the three SMDT scales, i.e., Melody, Rhythm, and Pitch are summarized in Table 1. Internal consistencies and split-half reliabilities were excellent (.79–.89) for all three scales.

Table 2 summarizes intercorrelations between the SMDT scales. Moderate positive correlations (r values .27–.41) were found between all SMDT scales (values above the diagonal). All correlations remained significant when controlling for total musical training (values below the diagonal).

All SMDT scales were associated with both musical experience, i.e. having played a musical instrument, as well as formal musical education. Table 3 summarizes group differences between individuals with and without musical experience (upper half) and between individuals with and without musical education (lower half). Individuals that had actively played a musical instrument showed higher scores than individuals that had never been musically active. Effect sizes (Cohen's d) ranged between .38 and .63. Likewise, individuals with musical education scored higher on the SMDT than participants that never had taken music lessons. The size of this effect was comparable to the effect of having played an instrument (Cohen's d .35–.60). A multiple regression model including both musical experience and musical education as independent variables only explained 0.6–1.2% more variance in musical ability than either experience or education alone. Among those participants that had played a musical instrument, there was a significant association between total hours of musical training and SMDT scores (Table 4). The size of these correlations ranged between .14 and .28 (Pearson r values). All reported effects were highly significant (p values <.0001).

Table 1
Descriptive statistics, internal consistency and reliability of the SMDT in the two subsamples.

	Sample 1	Sample 2
Melody		
Mean (SD)	6.70 (2.91)	6.67 (2.82)
Range (min–max)	0–18	0–18
Cronbach's alpha	.81	.79
Split-half coefficient	.82	.81
Rhythm		
Mean (SD)	15.36 (2.21)	15.29 (2.22)
Range (min–max)	4–18	4–18
Cronbach's alpha	.82	.81
Split-half coefficient	.83	.84
Pitch		
Mean (SD)	18.13 (4.85)	18.18 (4.83)
Range (min–max)	1–27	1–27
Cronbach's alpha	.87	.87
Split-half coefficient	.87	.89

Table 2
Intercorrelations between the SMDT scales. Values above the diagonal are *r* values for the raw Pearson correlations. Values below the diagonal are *r* values of partial correlations, controlled for total number of hours of practicing. In each cell, the upper value refers to Sample 1, and the lower value to Sample 2. All correlations were significant at $p < .001$.

	Melody	Rhythm	Pitch
Melody	–	.39 .39	.41 .37
Rhythm	.36 .35	–	.32 .32
Pitch	.35 .30	.28 .28	–

Table 3
SMDT scores, musical training, and musical education. SMDT scores for subgroups with and without active musical experience, and with and without musical education. Scores are means with SD in parentheses. Effect sizes (Cohen's *d*) and *t* statistics refer to comparisons (two sample *t* tests) between subgroups. In each cell, the upper values refers to Sample 1, and the lower value to Sample 2. All effects were significant at $p < .001$.

	N	Melody	Rhythm	Pitch
Musical experience				
Played an instrument	2439 2367	7.19 (2.96) 7.09 (2.88)	15.61 (2.09) 15.55 (2.09)	18.92 (4.78) 18.96 (4.77)
Never played	953 941	5.50 (2.36) 5.67 (2.37)	14.77 (2.37) 14.67 (2.43)	16.12 (4.44) 16.23 (4.40)
t Value		15.7 13.4	10.2 10.5	15.6 15.2
Cohen's *d*		.63 .54	.38 .39	.61 .60
Musical education				
Music lessons	1781 1718	7.41 (3.01) 7.33 (2.92)	15.73 (2.00) 15.70 (.201)	19.31 (4.71) 19.47 (4.41)
No music lessons	1611 1590	5.94 (2.57) 5.99 (2.54)	14.98 (2.34) 14.87 (2.36)	16.83 (4.68) 16.79 (4.56)
t Value		15.2 14.1	10.1 11.0	15.3 16.6
Cohen's *d*		.53 .49	.35 .38	.53 .60

Table 4
SMDT scores and hours of musical training. Correlations (Pearson *r* values) between SMDT scores and total hours of musical training, among those participants that had played a musical instrument. In each cell, the upper values refers to Sample 1, and the lower value to Sample 2. All correlations were significant at $p < .001$.

	Melody	Rhythm	Pitch
Total hours of musical training	.24 .27	.14 .17	.23 .28

3.2. Genetic modelling

Preliminary analyses showed that sex had a small mean effect on Pitch ($t(6715) = 6.98$, $p < 0.001$) with a slightly lower mean for females ($M = 17.80$, $SE = .07$) than for males ($M = 18.65$, $SE = .10$). Age showed a significant mean effect on Rhythm ($\beta = -.11$, $t(6878) = -9.21$, $p < .001$) and Pitch ($\beta = -.06$, $t(6715) = -5.06$, $p < .001$), with decreased discrimination skills with increased age. All genetic analyses were corrected for age and sex. Twin correlations and A, C and E estimates derived from univariate modelling are shown in Table 5. Twin correlations for Pitch and Melody suggested potential sex-limitation, with male DZ twins showing similar correlations to male MZ twins. Therefore, univariate general sex-limitation models were fitted first in order to explore sex-specific genetic influences. Sex-specific sources of genetic variance were non-significant, therefore common sex-limitation modeling was used subsequently, allowing the ACE estimates to differ quantitatively between the sexes. Results showed moderate heritability estimates for the Melody and Rhythm subtests of 59% and 50%, respectively, with no significant sex-limitation. The remaining variance was explained by non-shared environmental influences. Pitch, however, showed significant sex differences suggesting potential (though not significantly) higher heritability in females (30%) than in males (12% – non-significant) with additional shared environmental influences in males (38%) as opposed to only 19% in females (non-significant). Full modelling results not allowing for sex-differences in Pitch resulted in a heritability of 40% and non-significant shared environment (8%).

4. Discussion

We present a characterization of the reliability and validity of a new musicality test, the SMDT, in a large twin cohort. The data show that all three scales of the SMDT have good psychometric characteristics and a moderate heritability with some indication of potential sex-differences in their aetiology. In general, all test characteristics were highly similar in the two random split samples, further supporting the robustness of the findings across samples.

4.1. Reliability

Spearman–Brown split-half reliabilities and Cronbach's alpha values were in the range .79–.89 for all scales. While minimum recommendations for coefficient alpha vary between authors and are obviously to some extent a matter of subjective opinion, there appears to be broad agreement that alpha values above .8 indicate a high internal consistency of a scale (Nunnally & Bernstein, 1994; Streiner, 2010). In terms of internal consistency the SMDT scales also appear to fare well in comparison to other tests of musical perception. For the Profile of Music Perception Skills test Law and Zentner (2012) report Cronbach's alpha values between .48 and .78 for individual subscales, and between .85 and .87 for the composite scores. Seashore, Lewis, and Saetveit (1960) found internal consistencies (KR-21 coefficients, which are closely related to Cronbach's alpha) between .55 and .84 for the scales in the Seashore Measures of Musical Talent. Gordon reports split-half reliabilities in the range .51–.92 for the Seashore test and in the range .85–.93 for his own Music Aptitude Profile test (Gordon, 1969). Slightly lower values (.79–.85) were reported by Lee (1967) for Gordon's test. Impressively high Cronbach's alpha values have been reported for the subtests of the Musical Ear Test (Melody: .96; Rhythm: .94) (Wallentin et al., 2010). It should be noted, though, that these subtests consist of 52 items each, in contrast to the 18 items used for the corresponding scales in the present test. Cronbach's alpha increases with number of items, and some authors consider alphas >.9 as indicative of item redundancy (Streiner, 2010). Further details on reliabilities of various other musicality tests can be found elsewhere (Law & Zentner, 2012; Shuter-Dyson & Gabriel, 1981; Wallentin et al., 2010).

4.2. Validity

Criterion validity of the SMDT is supported by the fact that all subscales showed significant associations with both musical education and musical training. Participants with musical education showed higher mean scores than participants without musical education. Likewise, participants that actively had trained a musical instrument showed higher scores than participants who never

Table 5
ACE estimates for the three musical aptitudes (Rhythm, Melody, and Pitch) based on univariate modeling corrected for sex and age.

	Twin correlations (95% confidence intervals)		
	Rhythm	Melody	Pitch
MZf	0.52 (0.45; 0.58)	0.59 (0.53; 0.64)	0.45 (0.43; 0.52)
MZm	0.50 (0.39; 0.58)	0.53 (0.43; 0.60)	0.51 (0.41; 0.59)
DZf	0.27 (0.14; 0.39)	0.25 (0.13; 0.36)	0.33 (0.20; 0.44)
DZm	0.30 (0.12; 0.45)	0.45 (0.30; 0.56)	0.49 (0.36; 0.60)
DZos	0.28 (0.17; 0.38)	0.29 (0.18; 0.39)	0.17 (0.06; 0.27)
	ACE estimates (95% confidence intervals)		
	Rhythm	Melody	Pitch (females/males)*
A	0.50 (0.33; 0.57)	0.59 (0.43; 0.64)	0.30 (0.09; 0.52) / 0.12 (0.00; 0.35)
C	0.02 (0.00; 0.17)	0.01 (0.00; 0.14)	0.19 (0.00; 0.38) / 0.38 (0.17; 0.53)
E	0.48 (0.43; 0.53)	0.40 (0.36; 0.45)	0.52 (0.46; 0.58) / 0.50 (0.43; 0.59)

Abbreviations: A = additive genetic influences; C = shared environmental influences; E = non-shared environmental influences; f = females; m = males; MZ = monozygotic; DZ = dizygotic; os = opposite-sex.

* Pitch showed significant sex-differences.

had played. Furthermore, among the musically active participants there was a significant relation between SMDT scores and hours of musical training.

As expected, given the previous literature, all three SMDT scales showed moderate positive intercorrelations (Carroll, 1993; Law & Zentner, 2012; Wallentin et al., 2010). This can be taken as evidence for convergent validity of the scales in the sense that they all partly reflect an underlying, more general musicality factor. To some extent, these positive associations are also likely to reflect general intelligence, which is correlated with a broad range of sensory discrimination tasks (Lynn & Gault, 1986; Spearman, 1904; Troche & Rammsayer, 2009). It appears likely that stronger correlations would be found between the individual SMDT subscales and corresponding subtests of other auditory musical discrimination tests, e.g. the Seashore test, the Musical Ear Test, or the Profile of Music Perception Skills (Law & Zentner, 2012; Wallentin et al., 2010).

4.3. Twin modelling

The music abilities showed moderate genetic influences with slightly lower estimates (though non-significant) in males compared to females. In males only, there was an additional significant shared-environmental influence on Pitch, suggesting potential differences in the underlying etiology in music skills between sexes. To our knowledge, the only previous study examining heritability of musical discrimination was performed by Drayna, Manichaikul, de Lange, Snieder, and Spector (2001). In that study, a high heritability (80%) was found for the Distorted Tunes Test (DTT) in a sample of 284 all-female twin pairs. The DTT requires participants to determine whether melodic stimuli contain incorrect pitches or not. It thus corresponds most closely to the Melody subtest employed here. Notably, we observed the highest heritability for the Melody scale in females (63%). However, the melodic stimuli employed in the DTT were much longer (12–26 notes) than in the present test (5–9 notes). Conceivably, this implied a higher load on working memory and intelligence for the DTT, which could explain the higher heritability found by Drayna and coworkers. Genetic and environmental influences on the relationship between the three sub-tests as well as with IQ will be discussed elsewhere (Mosing et al. in preparation).

In summary, we hope that the SMDT will serve as a useful complement to existing tests of auditory musical discrimination. Notable features of the test are that it has a short test taking time (around 3 min for Rhythm and Pitch and ca 4.5 min for Melody), is adapted for online administration, and has a suitable difficulty level for general populations. The present findings demonstrate its reliability and validity in a large cohort of participants. Analyses of associations between SMDT scores and other variables, including both various measures of musical engagement and psychological traits, and their underlying genetic and non-genetic influences will be presented in forthcoming studies.

Finally, a few limitations of the study should be mentioned. First, the response rate was relatively low (ca 21% of the total cohort). Since this data collection was part of a large research project focussing on music related questions, it appears likely that the participants who chose to complete the study have a somewhat higher average degree of musical interest and ability than typical. The SMDT scores presented here are thus not necessarily representative of the general Swedish population. Secondly, while Internet testing offers excellent possibilities to acquire psychological measurements from large cohorts, an obvious limitation is increased method variance due to the fact that the participant complete the tests at home on their own schedule. For auditory discrimination tests, variability in quality of the presentation of the acoustical stimuli could potentially influence performance. However, given the good reliability and validity of the SMDT subtests, we do not believe this was a serious problem in the present study.

Acknowledgements

This work was supported by the Bank of Sweden Tercentenary Foundation (M11-0451:1) and the Sven and Dagmar Salén Foundation.

References

Boyle, J. D., & Radocy, R. E. (1987). *Measurement and evaluation of musical experiences*. New York: Schirmer Books.

Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. New York: Cambridge University Press.

Chin, T. C., & Rickard, N. S. (2012). The Music USE (MUSE) Questionnaire: An instrument to measure engagement in music. *Music Perception, 29*, 429–446.

Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., & Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science, 291*(5510), 1969–1972.

Gordon, E. (1969). Intercorrelations among musical aptitude profile and Seashore measures of musical talents subtests. *Journal of Research in Music Education, 17*, 263–271.

Helmbold, N., Troche, S., & Rammsayer, T. (2007). Processing of temporal and nontemporal information as predictors of psychometric intelligence: A structural-equation-modeling approach. *Journal of Personality, 75*(5), 985–1006.

Jacobs, C. (1960). Psychology of music: Some European studies. *Acta Psychologica, 17*, 273–297.

Law, L. N. C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLoS One, 7*(12), e52508. http://dx.doi.org/10.1371/journal.pone.0052508.

Lee, R. E. (1967). An investigation of the use of the musical aptitude profile with college and university freshman music students. *Journal of Research in Music Education, 15*, 278–288.

Lichtenstein, P., De Faire, U., Floderus, B., Svartengren, M., Svedberg, P., & Pedersen, N. L. (2002). The Swedish twin registry: A unique resource for clinical, epidemiological and genetic studies. *Journal of Internal Medicine, 252*(3), 184–205.

Lichtenstein, P., Sullivan, P. F., Cnattingius, S., Gatz, M., Johansson, S., Carlström, E., et al. (2006). The Swedish twin registry in the third Millenium: An update. *Twin Research and Human Genetics, 9*(6), 875–882.

Lynn, R., & Gault, A. (1986). The relation of musical ability to general intelligence and the major primaries. *Research in Education, 36*, 59–64.

Lynn, R., Wilson, R. G., & Gault, A. (1989). Simple musical tests as measures of Spearman's g. *Personality and Individual Differences, 10*(1), 25–28.

Magnusson, P. K. E., Almqvist, C., Rahman, I., Ganna, A., Viktorin, A., Walum, H., et al. (2013). The Swedish twin registry: Establishment of a biobank and other recent developments. *Twin Research and Human Genetics, 16*(1), 317–329.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2006). *Mx: Statistical modeling* (7th ed.). Richmond, USA: Department of Psychiatry.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Schulze, K., Dowling, W. J., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and backward recognition task. *Music Perception, 29*, 255–267.

Seashore, C. E. (1919). *The psychology of musical talent*. Boston, MA, USA: Silver, Burdett and company.

Seashore, C. E. (1938). *The psychology of music*. New York, NY: McGraw Hill.

Seashore, C. E. (1947). *In search of beauty in music*. New York: Ronald Press.

Seashore, C. E., Lewis, D., & Saetveit, J. G. (1960). *A second revision of the manual of instructions and interpretations for the seashore measures of musical talents (1939 revision)*. New York, NY: The Psychological Corporation.

Shuter-Dyson, R. (1999). Musical ability. In D. Deutsch (Ed.), *The psychology of music* (pp. 627–652). San Diego: Academic Press.

Shuter-Dyson, R., & Gabriel, C. (1981). *The psychology of musical ability*. London: Methuen.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15*, 201–293.

Streiner, D. L. (2010). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personal Assessment, 80*(1), 99–103.

Troche, S. J., & Rammsayer, T. H. (2009). Temporal and non-temporal sensory discrimination and their predictions of capacity- and speed-related aspects of psychometric intelligence. *Personality and Individual Differences, 47*(1), 52–57.

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The musical ear test, a new reliable test for measuring musical competence. *Learning and Individual Differences, 20*, 188–196.