Test–Retest Reliability and Minimal Detectable Change of Randomized Dichotic Digits in Learning-Disabled Children: Implications for Dichotic Listening Training

Mohammad Ebrahim Mahdavi*†
Akram Pourbakht*‡
Akram Parand§
Shohreh Jalaie¶

Abstract

Background: Evaluation of dichotic listening to digits is a common part of many studies for diagnosis and managing auditory processing disorders in children. Previous researchers have verified test–retest relative reliability of dichotic digits results in normal children and adults. However, detecting intervention-related changes in the ear scores after dichotic listening training requires information regarding trial-to-trial typical variation of individual ear scores that is estimated using indices of absolute reliability. Previous studies have not addressed absolute reliability of dichotic listening results.

Purpose: To compare the results of the Persian randomized dichotic digits test (PRDDT) and its relative and absolute indices of reliability between typical achieving (TA) and learning-disabled (LD) children.

Research Design: A repeated measures observational study.

Study Sample: Fifteen LD children were recruited from a previously performed study with age range of 7–12 yr. The control group consisted of 15 TA schoolchildren with age range of 8–11 yr.

Data Collection and Analysis: The Persian randomized dichotic digits test was administered on the children under free recall condition in two test sessions 7–12 days apart. We compared the average of the ear scores and ear advantage between TA and LD children. Relative indices of reliability included Pearson's correlation and intraclass correlation ($ICC_{2,1}$) coefficients and absolute reliability was evaluated by calculation of standard error of measurement (SEM) and minimal detectable change (MDC) using the raw ear scores.

Results: The Pearson correlation coefficient indicated that in both groups of children the ear scores of test and retest sessions were strongly and positively (greater than +0.8) correlated. The ear scores showed excellent ICC coefficient of consistency (0.78–0.82) and fair to excellent ICC coefficient of absolute agreement (0.62–0.74) in TA children and excellent ICC coefficients of consistency and absolute agreement in LD children (0.76–0.87). SEM and SEM% of the ear scores in TA children were 1.46 and 1.44% for the right ear and 4.68 and 5.47% for the left ear. SEM and SEM% of the ear scores in LD children were 4.55 and 5.88% for the right ear to 7.56 and 12.81% for the left ear. MDC and MDC% of the ear scores in TA children varied from 4.03 and 3.99% for the right ear to 12.93 and 15.13% for the left ear. MDC and MDC% of the ear scores in LD children varied from 12.57 and 16.25% for the right ear to 20.89 and 35.39% for the left ear.

Conclusions: The LD children indicated test–retest relative reliability as high as TA children in the ear scores measured by PRDDT. However, within-subject variations of the ear scores calculated by indices of absolute reliability were considerably higher in LD children versus TA children. The

---

*Department of Audiology, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran; †Department of Audiology, School of Rehabilitation, Shahid Beheshti University of Medical Sciences, Tehran, Iran; ‡Rehabilitation Research Center, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran; §Faculty of Psychology and Education, University of Tehran, Tehran, Iran; ¶Department of Physiotherapy, School of Rehabilitation, Tehran University of Medical Sciences, Tehran, Iran

Corresponding author: Akram Pourbakht, Department of Audiology, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran; Email: pourbakht.a@iums.ac.ir

results of the current study could have implications for detecting real training-related changes in the ear scores.

Key Words: dichotic listening, digits, learning disability, minimal detectable change, test–retest reliability

Abbreviations: CI = confidence interval; CV = consonant-vowel; DEA = dominant ear advantage; ICC = intraclass correlation; LD = learning disabled; LE = left ear; LEA = left-ear advantage; MDC = minimal detectable change; PRDDT = Persian randomized dichotic digits test; RDDT = randomized dichotic digits test; RE = right ear; REA = right-ear advantage; SD = standard deviation; SEM = standard error of measurement; TA = typical achieving

INTRODUCTION

Dichotic listening tests are among the most common central behavioral tests used to assess the functioning of the cerebral hemispheres, the interhemispheric transmission of information, the central auditory nervous system maturation, and central auditory processing disorders in adults and children (Keith and Anderson, 2007). Various materials are applicable in dichotic speech tests, including nonsense consonant-vowel (CV) syllables, digits, monosyllabic words, and sentences (Noffsinger et al, 1994; Jerger and Martin, 2006; Obrzut and Mahoney, 2011).

Digits represent suitable speech materials for assessment of dichotic listening in children with a wide range of linguistic ability. However, older children who are tested with a double dichotic digits test may show a ceiling effect (Strouse and Wilson, 1999a; Strouse et al, 2000; Moncrieff and Musiek, 2002; Neijenhuis et al, 2003). This led to the development of the American randomized dichotic digits test (Department of Veterans Affairs, 1998). The randomized dichotic digits test (RDDT) induces an uncertainty so that the listener does not know in advance the number of digit-pairs in the incoming dichotic item. This increases the difficulty of the test and reduces the performance of the both ears, more pronouncedly the performance of the left ear (Strouse and Wilson, 1999a). According to Strouse and Wilson (1999b), the two-pair component of the RDDT is challenging enough to be a sensitive indicator of binaural integration (Strouse and Wilson, 1999b; Moncrieff et al, 2016). Currently, RDDT normative data are available for adults (Strouse and Wilson, 1999b) and children (Moncrieff and Wilson, 2009; Moncrieff, 2011).

The RDDT has recently been adapted for use with the Persian language by Mahdavi et al (2015) in two lists (Mahdavi et al, 2015). Both the American and Persian RDDTs (PRDDTs) use nine monosyllabic digits 1–10 (except for disyllabic digit 7 in the American version and digit 4 in the PRDDT). In the Persian version of this test, similar to the original American test (Department of Veterans Affairs, 1998), there are 500-msec intervals between the digits and 4- to 8-sec intervals between the items so that 8 sec is given to the test participants to repeat the digits after the three-pair items, 6 sec after the two-pair items, and 4 sec after the one-pair items. Supplemental Audio S1

is list 1 of the PRDDT that is available with the online version of this article. Similar to its American counterpart, each list in the PRDDT contains 54 items that are equally divided into the one-, two-, and three-pair items. The RDDT has a unique structure with one-, two- and three-pair digits that are randomly distributed in the list. Each ear gets 108 scores when receiving the full raw score in each list so that 50% of this total score is obtained from the three-pair digits, 33% from the two-pair digits, and 17% from the one-pair digits. Therefore, the weight of the three-pair digits in the total score of the RDDT is higher than that of the two- and one-pair digits. Supplemental Appendix S1 is the score sheet of the PRDDT that is available with the online version of this article.

A majority of normal right-handed participants show a right-ear advantage (REA) in dichotic listening; this means that the number of correct answers for the right ear as the dominant ear for language is higher than those for the nondominant left ear (Moncrieff, 2011). According to the anatomical model of Kimura, observation of REA in right-handed individuals has its origin in the contralateral auditory pathway, which is a strong neural route for conduction of speech signals to the language brain area located within the left hemisphere. Auditory information received by the right ear will follow a shorter path to the language-dominant hemisphere. However, auditory information received by the left ear has to cross the corpus callosum, which may impose a lag relative to the information of the right ear. Therefore, the speech signal to the right ear is processed stronger and faster within the left hemisphere and suppresses the ipsilateral route of the speech signal to the left ear ascending to the left hemisphere (Moncrieff and Black, 2008; Musiek and Weihing, 2011). It seems that working memory plays a role in the size of REA. Experiments with dichotic digits have shown that when the number of digit pairs increases from one pair to three pairs, REA increases due to inefficiency of the left ear in dichotic listening (Wilson and Jaffe, 1996; Strouse and Wilson, 1999a; Moncrieff and Wilson, 2009). Penner et al (2009) highlighted the initial work of Kimura (1961; 1967) and showed that the magnitude of REA increases as the number of letters in a pair increases from 3 to 5 (Kimura, 1967; 1961; Penner et al, 2009). REA of the PRDDT in a group of 18- to 25-yr-old adults did not differ from REA of the three-pair dichotic digits (Aghazadeh

et al, 2015; Mahdavi et al, 2015). However, calculating interaural asymmetry in dichotic listening using the traditional method of REA has a disadvantage; individuals indicating left-ear advantage (LEA) yield negative values for REA that lessen the average interaural asymmetry of a population. Moncrieff (2011) recommends that ear advantage be calculated by subtracting the score of the nondominant ear from the dominant ear to compensate for the bias produced by the participants with LEA.

There are studies aimed at test–retest reliability of dichotic listening in children (Mukari et al, 2006), adults (Hugdahl and Hammar, 1997; Strouse and Wilson,1999b), and patients (Koomar and Cermak, 1981; Strouse and Hall, 1995). As part of a study by Mukari et al (2006), a 4-week test–retest reliability of Malay one- and two-paired dichotic digits was evaluated in normal children aged 6–11 yr in the free recall condition. The right- and left-ear scores in the retest session differed significantly from those in the test session and the average improvement of the scores in the retest session was 2.28 and 5.16% for the one-paired dichotic digits and 3.06 and 6.58% for the two-paired dichotic digits. In that study, ear advantage of the one-paired digits was decreased significantly in the retest session (Mukari et al, 2006). Strouse and Hall (1995) measured test–retest reliability of two-paired dichotic digits developed by Musiek (1983) in 10 adults with Alzheimer's disease. The control group consisted of ten normal adults. Pearson's correlation coefficients between the test and retest session in the Alzheimer's group were 0.82 and 0.97 for the right- and left-ear scores, respectively. The results of the Alzheimer's group were comparable to the control group ($r = 0.79$ and $r = 0.85$ for the right- and the left-ear scores, respectively), although variability of scores of the Alzheimer's group in the first session was higher than the control group. In that study, changes of the scores in retest session did not exceed 10% in any participant except for three cases of the Alzheimer's group with 11–13% score changes (Strouse and Hall, 1995). Strouse and Wilson (1999b) reported intertrial reliability of the American randomized dichotic digits test in adults aged 20–79 yr. The left-ear score of the 70- to 79-yr-old age group improved from 64.5% in the first trial to 70.2% in the second trial for three-pair digits (Strouse and Wilson, 1999b). Koomar and Cermak (1981) investigated a 1-week test–retest reliability of CV and three-pair digits materials in normal and learning-disabled (LD) children between the ages of 7 and 10 yr. Pearson's correlation and consistency of REA between test and retest sessions were considered as indices of test–retest reliability. The results showed that both groups showed an REA and that there were no significant differences in the size of ear advantage between the two groups of children on both dichotic CVs and digits. The normal children were more reliable than those in the LD group and both groups of the children

showed higher test–retest reliability in dichotic CVs than for the dichotic digits (Koomar and Cermak, 1981).

Relative reliability is defined as the degree to which individuals maintain their position in a sample with repeated measurements and usually is assessed by a form of correlation coefficient such as Pearson's correlation and intraclass correlation (ICC) coefficients. Absolute reliability is defined as the degree to which repeated measurements vary for studied individuals (Baumgartner, 1989). In contrast to relative reliability that evaluates association of two variables, absolute reliability is concerned with their proximity (Bruton et al, 2000). Indices of absolute reliability are expressed either in the actual units of measurement or as a proportion of the measured values (dimensionless ratio). Absolute reliability is estimated by indices such as standard error of measurement (SEM), method error, coefficient of variation, Bland–Altman limits of agreement, and minimal detectable change (MDC) (Atkinson and Nevill, 1998). Beckerman et al (2001) introduced the "smallest real difference" or MDC and defined it as the limit for the smallest change between two measurements, which indicates a real (clinical) change for a single individual following intervention (Beckerman et al, 2001).

Emerging documents suggest that dichotic listening training interventions such as Dichotic Interaural Intensity Difference and Auditory Rehabilitation for Interaural Asymmetry can ameliorate dichotic listening deficits (Musiek et al, 2004; 2008; Moncrieff and Wertz, 2008). We need indices for detecting the real change in the ear scores after a period of dichotic listening training. Thus, typical trial-to-trial variation of the ear scores is clinically important when measuring the outcome of a dichotic listening intervention.

Two previous studies showed that the PRDDT has enough difficulty to detect REA in young adults and has adequate test–retest relative reliability (Aghazadeh et al, 2015; Mahdavi et al, 2015). However, currently, there is no information about the test–retest reliability of the PRDDT in children. The studies on test–retest reliability of dichotic listening have been performed on normal children and adults using indices of the relative reliability (Bakker et al, 1978; Strouse and Hall, 1995; Hugdahl and Hammar, 1997; Mukari et al, 2006). The abovementioned studies have used Pearson's correlation coefficients as a relative indicator of test–retest reliability and average difference between the test and retest sessions to control systematic changes due to factors such as the learning effect. Learning disability is associated with dichotic listening deficit (Weihing and Musiek, 2013). Currently, there is no study on the absolute reliability of dichotic listening results in LD children as a target population for dichotic listening training. Therefore, this study was conducted with the purpose of comparing the results of the PRDDT and the relative and absolute indices of test–retest reliability between TA and LD children.

MATERIALS AND METHODS

This study was performed using a test–retest design in which the participants were tested with the PRDDT within a 7- to 12-day period.

Participants

Fifteen LD children (eight males and seven females) with an age range of 7–12 yr were recruited from a study performed by Esmaili et al (2016) on specific learning disability in Iran. The study psychologist (A. P.) confirmed the diagnosis of learning disability in the children based on the Wechsler intelligence scale for children, teacher-created tests based on school textbooks (Esmaili et al, 2016), and the definition of learning disability provided in Diagnostic and Statistical Manual of Mental Disorders-*IV* (American Psychiatric Association, 2000). Eleven children of the LD group were right handed and four of them were left handed. The control group consisted of 15 right-handed typical achieving (TA) children with an age range of 8–11 yr selected from three elementary schools whose first semester reports showed typical educational achievement. Hearing thresholds of both groups of the children were within normal limits ($\leq$15 dB HL in frequencies 500–4000 Hz) and interaural hearing threshold asymmetry was $\leq$10 dB. The parents of the children signed a written consent form and the children were paid for participating in our study.

Procedures

A calibrated laptop (Dell Inspiron 6400, Dublin, Ireland) attached to headphones (Philips SHM 6500/10) was used to administer the PRDDT. Before starting the test, we explained the objective of the test and the participants were instructed how to perform the test. Further, the participants got familiar with the test procedure through seven practice items including one-, two-, or three-pair digits before starting the test session. List 1 of the PRDDT was performed on the children at 70 dB HL under free recall conditions in a very quiet room. The retest session was held within 7–12 days with the average of 9.3 days after the test session.

Statistics

REA was calculated through the traditional method in which the left-ear score was subtracted from the right-ear score with two directions, positive and negative. Dominant ear advantage (DEA) was calculated by subtracting the nondominant ear score from the dominant ear score. The normality assumption of the data was checked by the Kolmogorov–Smirnov statistical test using the statistical software SPSS 21.0 (IBM SPSS Inc., Chicago, IL). The average of the ear scores was compared between the right and left ears using a paired *t* test. An independent *t* test was used for comparing the average of the REA and DEA ear scores between TA and LD children. Fisher's exact test was used for comparing the direction consistency of REA between TA and LD children. A 0.05 significance level was considered for all statistical tests.

Relative test–retest reliability of the ear scores and REA were estimated using Pearson's correlation coefficient and intraclass correlation coefficient model ($ICC_{2,1}$) using a single measure two-way mixed effects model with both consistency and absolute agreement definitions. A 95% confidence interval (CI) was constructed around $ICC_{2,1}$ point estimation for the ear scores and REA. Interpretation of ICCs was based on the known Fleiss' classification of ICC as follows: <0.4: poor; 0.4–0.75: fair to good; and >0.75: excellent (Fleiss, 1986).

SEM and SEM%—as absolute test–retest reliability indices—were calculated using the raw ear scores. The square root of within-subjects mean squares was used for calculating SEM ($SEM = \sqrt{WMS}$). SEM% was calculated through dividing SEM by the mean of the test and retest data ($\bar{m}$) and multiplying the result by 100 ($SEM\% = 100 \times \frac{SEM}{\bar{m}}$) (Downham et al, 2005).

The MDC was obtained by multiplying SEM by $\sqrt{2}$ and by 1.96, ($MDC = SEM \times \sqrt{2} \times 1.96$). MDC% was computed by dividing MDC by the mean of the test and retest scores and multiplying the result by 100 ($MDC\% = 100 \times \frac{MDC}{\bar{m}}$). A reference band for MDC was constructed by calculating the mean of the difference between the test and retest ($\bar{d}$) and creating $\bar{d} \pm MDC$ (Downham et al, 2005).

RESULTS

There was no significant difference in the age means; the TA and LD children had age means ($\pm$standard deviations [SDs]) of 9.2 $\pm$ 1.5 and 8.3 $\pm$ 1.3 yr old, respectively. The average hearing thresholds of right and left ears in the frequencies of 500–4000 Hz did not show significantly difference between TA and LD children and between the ears of each group of the children (p ranging from 0.069 to 0.615 for between-group and 0.063 to 0.564 for within-group interaural hearing threshold asymmetry). Table 1 indicates demographics and the individual ear scores in percent correct and REA in percent and we report the results in the text as mean (in %) $\pm$ SD.

Figure 1 is a scatter plot with a 45° line that illustrates agreement between the test and retest results of the ear scores and REAs for the studied children in percent. The more closely the scatter data points cluster around the 45° line, the greater the agreement is between the test and retest results. A 45° line illustrates perfect agreement.

The mean RE score was significantly better than that of the LE score in both groups of children [TA children

Table 1. Characteristics and Individual Data of Children with Means and Standard Deviations of Ear Scores and REA in the Test and Retest Sessions (in %)

TA Children

Participant	Sex/Handedness/Age (yr)	Test			Retest		
		RE	LE	REA	RE	LE	REA
1	M/R/8	98.15	89.81	8.34	98.15	92.59	5.56
2	M/R/12	94.44	67.59	26.85	96.30	70.37	25.93
3	M/R/7	84.26	75.00	9.26	90.74	77.78	12.96
4	M/R/11	87.96	62.96	25.00	92.59	73.15	19.44
5	M/R/8	92.59	81.48	11.11	95.37	97.22	-1.85
6	M/R/8	91.67	58.33	33.34	94.44	65.74	28.70
7	M/R/9	91.67	80.56	11.11	93.52	87.96	5.56
8	M/R/8	90.74	62.96	27.78	92.59	65.74	26.85
9	M/R/11	97.22	72.22	25.00	96.30	75.00	21.30
10	F/R/8	91.67	79.63	12.04	96.30	82.41	13.89
11	F/R/9	92.59	83.33	9.26	94.44	98.15	-3.71
12	F/R/8	90.74	83.33	7.41	90.74	86.11	4.63
13	F/R/11	94.44	84.26	10.18	96.30	74.07	22.23
14	F/R/10	95.37	85.19	10.18	97.22	87.96	9.26
15	F/R/10	93.52	83.33	10.19	95.37	90.74	4.63
Mean (SD)		**92.47 (3.5)**	**76.67 (9.6)**	**15.80 (8.9)**	**94.70 (2.3)**	**81.67 (10.8)**	**13.02 (10.6)**

LD Children

Participant	Sex/Handedness/Age (yr)	Test			Retest		
		RE	LE	REA	RE	LE	REA
1	F/R/7	75.00	32.41	42.59	72.22	48.15	24.07
2	M/R/8	82.41	62.04	20.37	82.41	70.37	12.04
3	F/R/7	52.78	54.63	-1.85	53.70	61.11	-7.41
4	F/R/7	71.30	28.70	42.60	66.67	47.22	19.45
5	F/L/8	61.11	94.44	-33.33	62.04	93.52	-31.48
6	M/R/9	80.56	62.96	17.60	86.11	43.52	42.59
7	F/R/7	64.81	29.63	35.18	70.37	42.59	27.78
8	M/R/8	71.30	41.67	29.63	75.00	54.63	20.37
9	M/R/9	78.70	61.11	17.59	73.15	70.37	2.78
10	F/R/11	69.44	40.74	28.70	71.30	46.30	25.00
11	F/L/9	80.56	62.96	17.60	82.41	67.59	14.82
12	M/R/8	40.74	23.15	17.59	53.70	45.37	8.33
13	M/L/7	59.26	53.70	5.56	71.30	59.26	12.04
14	M/L/9	84.26	55.56	28.70	87.04	53.70	33.34
15	M/R/11	77.78	62.04	15.74	91.67	70.37	21.30
Mean (SD)		**70.00 (12.3)**	**51.05 (18.6)**	**18.95 (19.0)**	**73.27 (11.4)**	**58.27 (14.2)**	**15.00 (17.7)**

test session: $t_{(14)} = 6.9$, $p = 0.000$, retest session: $t_{(14)} = 4.8$, $p = 0.000$; LD children test session: $t_{(14)} = 3.9$, $p < 0.005$, retest session: $t_{(14)} = 3.27$, $p < 0.01$]. Therefore, the PRDDT revealed a clear average REA in both groups of children. Table 1 contains individual data, mean, and SD of the ear scores and REA recorded in the test and retest sessions for the groups. Statistical analysis detected a significantly poorer performance in LD children versus TA children in the both test and retest sessions for RE scores [the test session: $t_{(28)} = 6.8$, $p < 0.001$, $d = 2.48$; the retest session: $t_{(28)} = 7.15$, $p = 0. 0.000$, $d = 2.6$] and LE scores [the test session: $t_{(28)} = 4.74$, $p < 0.001$, $d = 1.73$; the retest session: $t_{(28)} = 5.08$, $p < 0.001$, $d = 1.85$].

In TA children, average REA and DEA (in %) for the test session was the same (15.8 ± 8.9) because none of the TA children produced left-ear dominancy and the average REA in the retest session (13.02 ± 10.6) did not differ significantly from the average DEA of the retest session (13.77 ± 9.5) [$t_{(14)} = -13.82$, $p = 0.189$]. In LD children, the average REA of the test session (18.95 ± 19.0) and retest session (15.00 ± 17.7) was lower than the corresponding average DEA of the test session (23.64 ± 12.11) and retest session (20.19 ± 10.9), although the observed differences were not statistically significant [test session: $t_{(14)} = -1.05$, $p = 0.308$; retest session: $t_{(14)} = -2.22$, $p = 0.242$] so we presented only REA in Table 1.

Between-group comparison of the score changes (in %) showed that the mean of the retest change of RE score (2.22 ± 1.9) and LE score (5.00 ± 6.1) in the TA children did not differ significantly from the mean retest change of RE scores (3.27 ± 5.9) and LE scores (7.22 ± 9.9) in the LD children, [$t_{(28)} = 0.650$, $p = 0.521$ for RE scores; $t_{(28)} = 0.739$, $p = 0.466$ for LE scores], respectively. The mean REA of LD children in the test and retest sessions did not differ significantly from the mean of the TA children in the test and retest sessions, respectively [$t_{(28)} = -0.581$, $p = 0.566$ for test REA; $t_{(28)} = 0.370$, $p = 0.714$ for retest REA].

Comparison of average DEA in LD children in test and retest sessions (23.64 ± 12.1 and 20.18 ± 10.9, respectively) did not show statistically significant difference from the corresponding average DEA in TA children (15.80 ± 8.9 and 13.77 ± 9.5, respectively) [$t_{(28)} = -2.02$, $p = 0.053$ for the test DEA; $t_{(28)} = -1.71$, $p = 0.097$ for the retest DEA]. Although average DEA in LD children was not statistically larger than average DEA in TA children, given the small number of participants and the medium effect size of this difference in the test (Cohen's $d = 0.74$) and retest sessions (Cohen's $d = 0.63$), the observed difference is clinically important. Since LEA is more prevalent in LD children (Keith, 2007), it is logical that we expect DEA to produce a larger interaural asymmetry than REA in this group of children.

As presented in Table 1, the average of the both ear scores in TA children enhanced significantly in the
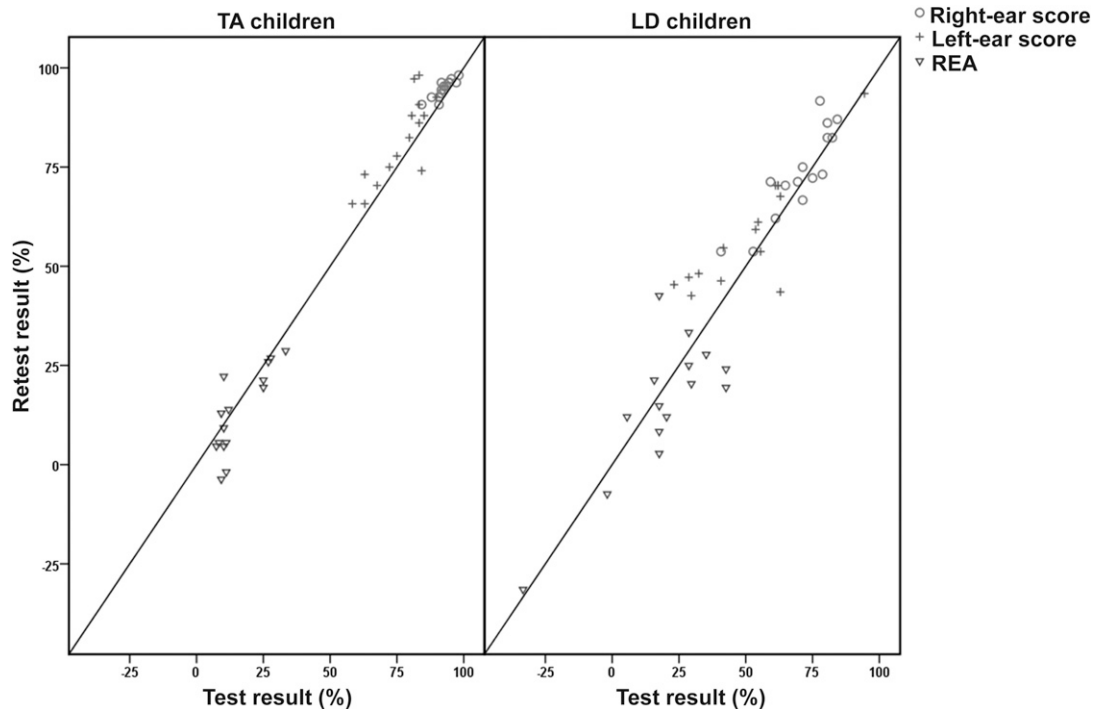
Figure 1. A bivariate plot showing test results (in %) on horizontal and retest results (in %) on vertical axis. The 45° diagonal line represents perfect agreement between the test and retest results.

retest session ($p < 0.001$ for RE, $p < 0.01$ for LE scores). In LD children, the RE score did not show a significant improvement, however, their average LE score significantly increased in the retest session ($p < 0.05$).

The average of the test REA of both groups of the children remained unchanged statistically in the retest session (Table 1). Thirteen of the TA children (86.7%) maintained the REA direction in the retest session and REA changed to LEA in two (13.3%) of the TA children. All of the LD children (100%) consistently indicated the same REA direction on the retest session. This difference was not statistically significant (p value calculated by one-sided Fisher's exact test = 0.241).

Table 2 contains ICC coefficients of consistency and agreement and Pearson's correlation coefficient between test and retest sessions for right and left scores and REA. The ICC coefficient of consistency of ear scores and REA was categorized as excellent (ICC = 0.78–0.87) in both groups of children. The ICC coefficient of agreement of the ear scores and REAs was categorized as fair to excellent in TA children and excellent in LD children. Similarly, as Pearson's correlation coefficient demonstrates, the test results were very strongly ($r > 0.80$) and positively correlated to retest results in both groups of the children. RE of the TA children obtained the lowest ICC coefficient of consistency and agreement (Table 2).

Table 2 presents SEM and SEM% for the ear scores and REAs. In both groups of children, SEM of the RE was lower than SEM of the left ear. The SEM value of REA in the LD children was higher than the corresponding value in the TA children. Calculated SEM% in LD children is 1.5–4 times higher than SEM% in TA children. Although ICCs and Pearson's correlation coefficient show a similar relative reliability of the ear scores and REAs for both groups of children, absolute reliability of the ear scores and REAs is considerably poorer in LD children.

Also presented in Table 2 are MDC and MDC%. In both groups of children, MDC of the right ear was lower than MDC of the left ear. The MDC value of REA in the LD children was higher than the corresponding value in the TA children.

DISCUSSION

The main purpose of the current study was evaluating and comparing the test–retest reliability of the PRDDT between TA and LD children using both indices of relative and absolute reliability and determining MDC for it. The results of this study showed that LD children performed more poorly in the PRDDT for both right and left ears than TA children. Both groups of children produced higher RE than LE scores but there was no significant difference between the average of REA in the LD and TA children. This is due to poorer performance of both ears of the LD children for the PRDDT. The results demonstrated that the PRDDT could detect a group difference between the TA and LD children only based on the ear scores.

Four LD children have LEA and we expected average DEA to be larger in size than average REA in these

Table 2. ICC Coefficients of Consistency and Agreement with 95% CI and Pearson's *r* with Calculated SEM, SEM%, MDC, MDC%, and Reference Band of 95% MDC for Ear Scores and REAs between Test and Retest Sessions in TA and LD Children

| | | ICC Coefficient (95% CI) | | Pearson's *r*** | SEM | SEM% | MDC | MDC% | Reference Band (95% MDC) |
		Consistency***	Agreement***						
TA children	Right ear	0.78 (0.47 to 0.92)	0.62 (−0.05 to 0.88)	0.86	1.46	1.44	4.03	3.99	−1.60 to 6.43
	Left ear	0.82 (0.57 to 0.94)	0.74 (0.22 to 0.92)	0.83	4.68	5.47	12.93	15.13	−7.53 to 17.33
	REA	0.80 (0.51 to 0.93)	0.78 (0.46 to 0.92)	0.81	4.34	27.87	11.99	77.03	−14.99 to 8.99
LD children	Right ear	0.87 (0.67 to 0.96)	0.85 (0.57 to 0.95)	0.88	4.55	5.88	12.57	16.25	−9.04 to 16.10
	Left ear	0.82 (0.55 to 0.94)	0.76 (0.30 to 0.92)	0.85	7.56	12.81	20.89	35.39	−13.09 to 28.69
	REA	0.80 (0.49 to 0.93)	0.79 (0.49 to 0.92)	0.81	8.31	45.34	22.97	125.29	−27.24 to 18.70

Notes: ***$p < 0.001$ for the ear scores and REA; **$p < 0.01$ for the ear scores and REA.

children but two methods (REA versus DEA) of calculating interaural asymmetry of dichotic listening did not show statistically significant average difference in LD children or TA children. However, Moncrieff (2011) demonstrated that DEA detected a larger average interaural asymmetry than REA in normal children. The discrepancy might be due to a difference in scoring technique, which is based on two-pair components of the RDDT in the American format and overall scores of all items in the Persian RDDT. More research with similar inclusion criteria, sample size, and scoring is needed for comparing the American and Persian RDDT.

We did not find any study using total-score RDDT with learning disability for comparison. Most of the studies on learning disability with digit materials have used double dichotic digits (Moncrieff and Musiek, 2002; Moncrieff and Black, 2008; Pinheiro et al, 2010; Ghannoum et al, 2014) that are easier than RDDT. However, our results are consistent with many studies summarized by Moncrieff et al (2016) including out-of-date studies (Thomson, 1976; Keefe and Swinney, 1979; Pelham, 1979; Aylward, 1984) on children with a reading disability—a specific learning disability—and a recent study (Pinheiro et al, 2010) on LD children.

Reliability Indices

According to Downham et al (2005), the reliability cannot be analyzed based only on relative indicators and an analysis of measurement errors must be performed as a complement (Downham et al, 2005). Since each of the reliability indices has their own advantages and disadvantages, researchers emphasize considering both absolute and relative reliability coefficients (Lexell and Downham, 2005).

Mean Difference

The mean difference between test and retest sessions demonstrated that the ear scores were susceptible to a learning effect. In the TA children the scores of both ears

and in the LD children the scores of the left ear improved in the retest session. LD children usually show a LE deficit in dichotic listening but some also demonstrate a RE deficit. Both deficits are targeted by auditory rehabilitation for interaural asymmetry (Moncrieff and Wertz, 2008). Because dichotic listening training programs seek to diminish this weakness, information about the learning effect of RDDT is clinically important. The LD children without any intervention increased their average LE scores significantly, by 7.22%. A similar improvement (7%) was reported by Weihing and Musiek (2013) for the weaker ear of a control group of children when tested pre- and post-dichotic interaural intensity difference training by double dichotic digits test (Weihing and Musiek, 2013). This index of reliability does not provide information regarding individual differences and it is better to complement it with other reliability indices (Weir, 2005; Zaki et al, 2013).

Pearson's Correlation Coefficient

In both groups of children, Pearson's *r* showed a very strong positive relationship between the results of test and retest sessions. Pearson's *r* for the ear scores obtained in the current study was in line with the study by Strouse and Hall (1995) and consistent with the upper range of correlations (0.6–0.8) reported in the previous studies using CV materials (Hugdahl and Hammar, 1997; Gadea et al, 2000). Pearson's *r* shows how test and retest results vary together without any information about the agreement of the results (Bruton et al, 2000). A strong correlation between the two sets of data does not guarantee that the two sets are highly repeatable, as Pearson's *r* does not see systematic or fixed errors. Thus, this index should be used in conjunction with other indicators for assessment of reliability (Weir, 2005; Zaki et al, 2013).

ICC: Consistency and Agreement

In the current study, both consistency and agreement of dichotic listening performance were calculated using the ICC method. The ICC model used in the current

study is similar to Pearson's r. As shown in Table 2, the ICCs with the consistency definition and Pearson's r coefficients are equivalent. Indices of agreement between the two measurements such as dichotic listening scores before and after dichotic listening training could be used for specifying expected improvement after intervention (de Vet et al, 2006). The results demonstrated that absolute agreement between test and retest scores was lower than the consistency of the scores. This discrepancy originates from the fact that ICC agreement is defined as the extent to which identical measurements are obtained in test and retest sessions and considers both systematic and random errors while ICC consistency does not take into account systematic differences (learning or practice effect) in the scores between the test and retest sessions (Downham et al, 2005; de Vet et al, 2006). As Figure 1 displays, the RE scores of the TA children have the lowest dispersion and are very close to the perfect agreement line. However, its ICC coefficient unexpectedly is the lowest in comparison to other variables. This is due to the dependency of ICC to heterogeneity of studied participants in performance measured. If between-subjects variability is low but real trial-to-trial consistency of a measured performance is high, it may result in a small ICC (Weir, 2005). This relatively misleading result of ICC has occurred for the RE scores of the TA children because TA children are less heterogeneous for the RE score than the LE score, REA, and the results of LD children (Figure 1).

SEM and SEM%

Absolute reliability indicators such as SEM provide information about the degree by which the repeated measurements vary for individuals. Since SEM estimates random variation of the performance across repeated measures, it is an important indicator for test sensitivity in detecting a change of performance over time. If SEM of the performance in a test–retest study is small, detection of intervention-related changes is easier (Downham et al, 2005; Lexell and Downham, 2005). In a repeated measures design, it is not known how much of the variability originates from the changes in mean and how much from typical variation of measured performance. This limitation of SEM can be overcome by calculating SEM%, which expresses measurement variability as a coefficient of variation (Lexell and Downham, 2005).

The previous studies on reliability of dichotic listening have not addressed within-subject variation of the ear scores or REAs. SEM indicates the precision of the ear scores and a CI can be constructed based on its value to identify true ear scores (Downham et al, 2005). SEM% specifies how much of the observed changes in the ear scores and REAs can be attributed to typical variation. This type of reliability may be clinically useful when

changes of the ear scores are detected after a period of dichotic listening training.

As presented in Table 1, the average REAs in both groups of the TA and LD children did not show statistically significant differences between test and retest sessions. This may be interpreted as high reliability of REAs. However, comparing the averages between the two sets of measurements did not provide any information on individual differences (Bruton et al, 2000). On the other hand, SEM% of REAs is considerably greater than that of the ear scores (Table 2). Typical variation of REA of the PRDDT in our sample of the LD children is 9.3 (45.34% × 20.47 row scores). High within-subject variation of REA in repeated measures may limit its adequacy as an indicator of posttraining improvement in LD children. However, the REA direction in TA and LD children showed a high intersession consistency. Previous studies have reported both consistency and inconsistency of REA in some of their participants. The more relevant study is Strouse and Wilson (1999b), which reported intertrial reversal of REA direction in five participants (25%) under 30 yr old for the American RDDT. Similarly, relative frequency of reversal of REA direction was 30% in a study on one-pair dichotic digits performed by Pizzamiglio et al (1974) and 29% in a study on CV syllables performed by Blumstein et al (1975). Koomar and Cermak (1981) found lower test–retest reliability of REA for three-pair dichotic digits in LD versus normal children with an age range of 7–10 yr. It seems that consistency of REA direction is also dependent on size of REA so that individuals with small REA are more likely to shift REA direction in the retest sessions (Blumstein et al, 1975). In the current study, all LD children maintained REA direction in the retest session while this consistency was observed in 86.7% of TA children with no significant difference between TA and LD children.

MDC and MDC%

MDC can be used to assess minimal changes of the ear scores required to be 95% confident that the induced changes after dichotic listening training are true changes and not measurement errors (Downham et al, 2005; Lexell and Downham, 2005). MDC and MDC% were higher in the LD children versus the TA children, especially for REA (Table 2).

To ensure that the dichotic listening deficit of a child measured by RDDT has clinically been changed after dichotic training, we need a "reference band" for RDDT. This range extracted from test and retest sessions was computed and presented in Table 2 and is entitled 95% MDC. If changes of ear scores after dichotic training are within this range, it cannot be considered a clinically important change. Since dichotic training is expected to increase ear scores, the improvement (posttreatment

score minus pretreatment score) should exceed the higher fence of this reference range. MDC%, which is independent of the units of measurement, is more easily interpreted. Suppose that we administer a dichotic training protocol on this group of LD children and expect to improve the left-ear deficit compared to pretreatment ear scores with a raw score of 55, SEM of 7.56, and SEM% of 12.81%. This means that the LE scores have a typical variation of 7.04 (12.81% × 55). We are 95% confident that the true raw score is within (55 ± [2 × 7.56]) or 39.8–70.1. If the posttraining score reaches 68, the audiologist cannot be sure that a true change has occurred. MDC%, presented in Table 2, implies that if we want to achieve a true change, the average of the LE raw score of the LD children has to exceed 74.4 (35.39% × 55 + 55).

As denoted in Table 2, in both groups of the children, MDC% of the ear scores is considerably lower than that of REA. This is because of the fact that REA is a difference value and is affected by typical variation of both the RE and LE scores. Therefore, considering indices of measurement variability for dichotic listening scores may facilitate clinical decision-making (Stratford, 2004).

The studies performed on dichotic listening training have not reported MDC and compared the outcome between experimental and control groups or used a before–after design for demonstrating efficacy of dichotic listening training (Katz et al, 1984; English et al, 2003; Moncrieff and Wertz, 2008). This study introduces MDC, which may be applicable for identifying real alternations of dichotic listening scores after dichotic listening interventions. However, according to Beckerman et al (2001) MDC is a clinometric indicator of the measuring tool and reflects the magnitude of change that confidently did not result from typical variation of the measured performance, whereas "clinically relevant change" is a change arbitrarily considered as important change by clinicians (Beckerman et al, 2001).

CONCLUSION

L D children showed test–retest relative reliability as high as TA children in the ear scores and in size and direction of REA measured by the PRDDT. However, the indices of absolute reliability revealed that the ear scores and REA were less reliable within the LD children versus the TA children. Establishment of a reference band of minimal detectable change may be useful for clinical tracking of training-related improvements in the ear scores.

REFERENCES

Aghazadeh J, Mahdavi ME, Tahaei S, Tabatabaee S. (2015) Inter-list equivalency and reliability of the Persian randomized dichotic digits test. *Aud Ves Res* 24:71–79.

American Psychiatric Association. (2000) *Diagnostic and Statistical Manual of Mental Disorders (DSM)*. Washington, DC: American Psychiatric Association.

Atkinson G, Nevill AM. (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 26(4):217–238.

Aylward EH. (1984) Lateral asymmetry in subgroups of dyslexic children. *Brain Lang* 22(2):221–231.

Bakker DJ, Van der Vlugt H, Claushuis M. (1978) The reliability of dichotic ear asymmetry in normal children. *Neuropsychologia* 16(6):753–757.

Baumgartner TA. (1989) Norm-referenced measurement: reliability. In: Safrit MJ, Wood TM, eds. *Measurement Concepts in Physical Education and Exercise Science*. Champaign, IL: Human Kinetics.

Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. (2001) Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res* 10(7):571–578.

Blumstein S, Goodglass H, Tartter V. (1975) The reliability of ear advantage in dichotic listening. *Brain Lang* 2(2):226–236.

Bruton A, Conway JH, Holgate ST. (2000) Reliability: what is it, and how is it measured? *Physiotherapy* 86:94–99.

de Vet HC, Terwee CB, Knol DL, Bouter LM. (2006) When to use agreement versus reliability measures. *J Clin Epidemiol* 59(10): 1033–1039.

Department of Veterans Affairs. (1998) *Materials for Auditory Perceptual AssessmentTonal and Speech Disc 2.0., VA Medical Center*. Mountain Home, TN: Rehabilitation Research and Development Service, VA Medical Center.

Downham DY, Holmba MA, Lexell J. (2005) Reliability of measurements in medical research and clinical practice. In: Paton R, McNamara L, eds. *Multidisciplinary Approaches to Theory in Medicine*. Amsterdam, the Netherlands: Elsevier. doi: 10.1016/S1571-0831(06)80013-4

English K, Martonik J, Moir L. (2003) An auditory training technique to improve dichotic listening. *Hear J* 56:34–36.

Esmaili SK, Shafaroodi N, Mehraban AH, Parand A, Qorbani M, Yazdani F, Mahmoudpour A. (2016) Prevalence of psychiatric symptoms and mental health services in students with specific learning disabilities in Tehran, Iran. *Int J Ment Health Addict* 14:438–448.

Fleiss JL. (1986) *Design and Analysis of Clinical Experiments*. 1st ed. New York, NY: John Wiley & Sons.

Gadea M, Gomez C, Espert R. (2000) Test-retest performance for the consonant-vowel dichotic listening test with and without attentional manipulations. *J Clin Exp Neuropsychol* 22(6):793–803.

Ghannoum MT, Shalaby AA, Dabbous AO, Abd-El-Raouf ER, Abd-El-Hady HS. (2014) Central auditory processing functions in learning disabled children assessed by behavioural tests. *Hear Balance Commun* 12:143–154.

Hugdahl K, Hammar A. (1997) Test-retest reliability for the consonant-vowel syllables dichotic listening paradigm. *J Clin Exp Neuropsychol* 19(5):667–675.

Jerger J, Martin J. (2006) Dichotic listening tests in the assessment of auditory processing disorders. *Audiol Med* 4:25–34.

Katz J, Chertoff M, Sawusch JR. (1984) Dichotic training. *J Aud Res* 24(4):251–264.

Keefe B, Swinney D. (1979) On the relationship of hemispheric specialization and developmental dyslexia. *Cortex* 15(3):471–481.

Keith RW. (2007) Diagnosing central auditory processing disorders in children. In: Roeser RJ, Valente M, Hosford-Dunn H, eds. *Audiology: Diagnosis*. New York, NY: Thieme Medical Publisher.

Keith RW, Anderson J. (2007) Dichotic listening tests. In: Museik F, Chermak GD, eds. *Handbook of (Central) Auditory Processing Disorder*. San Diego, CA: Plural Publishing.

Kimura D. (1961) Cerebral dominance and the perception of verbal stimuli. *Can J Psychol* 15:166.

Kimura D. (1967) Functional asymmetry of the brain in dichotic listening. *Cortex* 3:163–178.

Koomar JA, Cermak SA. (1981) Reliability of dichotic listening using two stimulus formats with normal and learning-disabled children. *Am J Occup Ther* 35(7):456–463.

Lexell JE, Downham DY. (2005) How to assess the reliability of measurements in rehabilitation. *Am J Phys Med Rehabil* 84(9):719–723.

Mahdavi ME, Aghazadeh J, Tahaei SAA, Heiran F, Akbarzadeh Baghban A. (2015) Persian randomized dichotic digits test: development and dichotic listening performance in young adults (in Persian). *Audiology* 23:99–113.

Moncrieff DW. (2011) Dichotic listening in children: age-related changes in direction and magnitude of ear advantage. *Brain Cogn* 76(2):316–322.

Moncrieff DW, Black JR. (2008) Dichotic listening deficits in children with dyslexia. *Dyslexia* 14(1):54–75.

Moncrieff D, Keith W, Abramson M, Swann A. (2016) Diagnosis of amblyaudia in children referred for auditory processing assessment. *Int J Audiol* 55(6):333–345.

Moncrieff DW, Musiek FE. (2002) Interaural asymmetries revealed by dichotic listening tests in normal and dyslexic children. *J Am Acad Audiol* 13(8):428–437.

Moncrieff DW, Wertz D. (2008) Auditory rehabilitation for interaural asymmetry: preliminary evidence of improved dichotic listening performance following intensive training. *Int J Audiol* 47(2):84–97.

Moncrieff DW, Wilson RH. (2009) Recognition of randomly presented one-, two-, and three-pair dichotic digits by children and young adults. *J Am Acad Audiol* 20(1):58–70.

Mukari SZ, Keith RW, Tharpe AM, Johnson CD. (2006) Development and standardization of single and double dichotic digit tests in the Malay language. *Int J Audiol* 45(6):344–352.

Musiek FE. (1983) Assessment of central auditory dysfunction: the dichotic digit test revisited. *Ear Hear* 4:79–83.

Musiek FE, Baran JA, Shinn J. (2004) Assessment and remediation of an auditory processing disorder associated with head trauma. *J Am Acad Audiol* 15(2):117–132.

Musiek FE, Weihing J. (2011) Perspectives on dichotic listening and the corpus callosum. *Brain Cogn* 76(2):225–232.

Musiek F, Weihing JA, Lau C. (2008) Dichotic interaural intensity difference (DIID) training: a review of existing research and future directions. *J Acad Rehabil Audiol* 41:51–65.

Neijenhuis K, Snik A, van den Broek P. (2003) Auditory processing disorders in adults and children: evaluation of a test battery. *Int J Audiol* 42(7):391–400.

Noffsinger D, Martinez CD, Wilson RH. (1994) Dichotic listening to speech: background and preliminary data for digits, sentences, and nonsense syllables. *J Am Acad Audiol* 5(4):248–254.

Obrzut JE, Mahoney EB. (2011) Use of the dichotic listening technique with learning disabilities. *Brain Cogn* 76(2):323–331.

Pelham WE. (1979) Selective attention deficits in poor readers? Dichotic listening, speeded classification, and auditory and visual central and incidental learning tasks. *Child Dev* 50(4):1050–1061.

Penner IK, Schläfli K, Opwis K, Hugdahl K. (2009) The role of working memory in dichotic-listening studies of auditory laterality. *J Clin Exp Neuropsychol* 31(8):959–966.

Pinheiro FH, Oliveira AM, Cardoso ACV, Capellini SA. (2010) Dichotic listening tests in students with learning disabilities. *Rev Bras Otorrinolaringol (Engl Ed)* 76(2):257–262.

Pizzamiglio L, De Pascalis C, Vignati A. (1974) Stability of dichotic listening test. *Cortex* 10(2):203–205.

Stratford PW. (2004) Getting more from the literature: estimating the standard error of measurement from reliability studies. *Physiother Can* 56:27–30.

Strouse AL, Hall JW, 3rd. (1995) Test-retest reliability of a dichotic digits test for assessing central auditory function in Alzheimer's disease. *Audiology* 34(2):85–90.

Strouse A, Wilson RH. (1999a) Stimulus length uncertainty with dichotic digit recognition. *J Am Acad Audiol* 10(4):219–229.

Strouse A, Wilson RH. (1999b) Recognition of one-, two-, and three-pair dichotic digits under free and directed recall. *J Am Acad Audiol* 10(10):557–571.

Strouse A, Wilson RH, Brush N. (2000) Recognition of dichotic digits under pre-cued and post-cued response conditions in young and elderly listeners. *Br J Audiol* 34(3):141–151.

Thomson ME. (1976) A comparison of laterality effects in dyslexics and controls using verbal dichotic listening tasks. *Neuropsychologia* 14(2):243–246.

Weihing J, Musiek FE. (2013) Dichotic interaural intensity difference (DIID) training. In: Chermak GD, Musiek FE, eds. *Handbook of Central Auditory Processing Disorder: Comprehensive Intervention*. San Diego, CA: Plural Publishing.

Weir JP. (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 19(1):231–240.

Wilson RH, Jaffe MS. (1996) Interactions of age, ear, and stimulus complexity on dichotic digit recognition. *J Am Acad Audiol* 7(5):358–364.

Zaki R, Bulgiba A, Nordin N, Azina Ismail N. (2013) A systematic review of statistical methods used to test for reliability of medical instruments measuring continuous variables. *Iran J Basic Med Sci* 16(6):803–807.

Supplemental Appendix S1

				Persian Randomized Dichotic Digits Test					
				List 1					
Score	**L**	**Score**	**R**		**Score**	**L**	**Score**	**R**	
	10_7		6_9	28		6_2_7		8_5_9	1
	10_6		3_7	29		2_1		3_9	2
	6_7_5		1_10_8	30		5		6	3
	10_7_2		6_9_5	31		10		6	4
	1		2	32		5_9_6		1_3_10	5
	8		5	33		6		7	6
	6_10_7		3_1_2	34		7_8		1_9	7
	7_1_3		2_8_9	35		7_8		5_6	8
	1_2		9_7	36		8_3_6		10_7_9	9
	2_1		1_8	37		10_3		9_8	10
	10		9	38		2_1_10		6_3_9	11
	10		2	39		5		2	12
	10_6_8		7_1_2	40		7_1_3		9_6_10	13
	8		9	41		1_6_5		8_3_7	14
	10_6		7_1	42		10_6_9		8_5_7	15
	7		10	43		9_10_8		7_5_3	16
	3		10	44		3_1_2		10_7_8	17
	2_5_10		8_3_7	45		7_5_9		3_2_1	18
	1		3	46		5		10	19
	6		8	47		3_7_10		9_8_5	20
	8_7		9_6	78		5		7	21
	2		8	49		2_8		9_1	22
	1_9		3_2	50		5_9		1_3	23
	8_2		5_1	51		3_9		8_7	24
	9_1		7_5	52		1		5	25
	10		1	53		7_1		9_6	26
	8_3		7_2	54		6_8_7		2_5_3	27
				Sum					**Sum**