

Fallacious Gender Differences on the Projective Drawing Test for Assessment of Emotional Behavioral Problems among Elementary School Children*

Sunghee An[†]

Hongik University

This study utilized projective drawing tests to identify whether the gender difference on a projective drawing test reflects actual differences of their psychological conditions between boys and girls. Despite their lack of objective scoring criteria, drawing tests have revealed that boys are more problematic both emotionally and behaviorally than girls. In this study, gender differences were examined via the Emotional-Behavioral Drawing Test. Six evaluators score the psychological conditions of 172 elementary school children. Positive automatic thought and emotional reactivity were measured as covariates. Results indicated there were statistically significant gender differences in drawing scores, but there were no significant effects of covariates. Given these results, it is plausible that no significant differences in psychological conditions existed between genders, even though boys may draw more psychologically problematic drawings than girls. It is suggested to apply gender criterion to score all types of drawing tests because boys and girls are different in drawings and expressing emotions.

Key words : gender differences, projective drawing tests, emotional-behavioral problems

* This work was supported by 2017 Hongik University Research Fund.

† Corresponding author: Sunghee An, Assistant Professor, Hongik University, Mapo-gu, Wausan-ro 94, Seoul, South Korea 04066, Tel : 02-320-3352, E-mail : ansh@hongik.ac.kr

Given the long-lasting belief that children's drawings are thought to be some expression of their psychological conditions, projective drawing tests have been frequently used by psychological counselors to check psychological difficulties without careful consideration of the appropriate scoring criterion and degree to which drawing data are supported by criterion-related validity psychometric testing results. Ever since various kinds of projective drawing tests were developed, researchers have repeatedly reported gender differences in children's drawings. With persistent arguments about gender differences in psychology due to biological differences and socio-cultural role differences (Blakemore, Berenbaum, & Liben, 2009), it had been asserted that the drawings of boys and girls should have been evaluated with separate norms from their early stages, since the beginning of the history of drawing tests (Goodenough, 1926; Harris, 1963). Though children's drawings have been used by professional counselors in clinical settings, sometimes used by non-professionals without sufficient training on interpretation of drawings in non-clinical settings such as schools for brief estimating cognitive development (Knoff & Prout, 1985; Naglieri, 1988) or emotional conditions such as attachment representations (Howard et al., 2017). It is therefore important to establish a reliable baseline of their properties, particularly regarding gender.

Generally speaking, projective drawing tests are evaluated based on contents of drawings

such as mood status of individuals and structure aspects such as coverage or graphic fluency. Most of criteria are advantageous for girls (Burns, 1987; Drake, Lo, Hwang, & Shin, 1995; Groth-Marnat & Roberts, 1998; Marzolf & Kirchner, 1973; Palmer et al., 2000; Zalsman et al., 2000). Researchers found that girls tend to be more detail-oriented than boys, having overall superiority in their drawings. Girls use double lines earlier than boys when they draw human limbs (Willsdon, 1977); and include more body parts and clothing in their figures than boys (Koppitz, 1968). Girls tend to use more colors overall in their free drawings (Turgeon, 2008). In family drawings, girls are more likely than boys to draw things reflecting their experiences with family relationships (Cherney, Seiwert, Dickey, & Flichtbeil, 2006).

In terms of content of drawings, boys tend to depict assaultive and stressful relationships in their drawings as compared to girls, who tend to portray friendly and caring relationships (Silver, 1993); boys tend to draw dynamic objects such as cars, machines, or robots whereas girls tend to draw nature and life-oriented objects such as humans, flowers, or butterflies (Cherney & London, 2006); girl drawings show inflexible positions of the face or body and show no action whereas boys' drawings depict more actions, such as doing sports (Malchiodi, 1990). Due to girls' tendency to value relationships more than boys; they are more likely to pay attention to what other same-sex friends wear

(Matlin, 2004); and girls draw more gender-typical content than boys do (Boyatzis & Eades, 1999). Girls draw more happy faces on people and animals including smiling suns or trees (Arteche & Murray, 2011).

Girls may have greater advantage than boys with the same criterion due to their better fine motor skills when their drawings are evaluated based on structure criteria (Halpern, 2004). Drawings by girls that have additional details could be misinterpreted as obvious superior intelligence because drawings are often interpreted based on the realistic completeness of figures (Pianta, Longmaid, & Ferguson, 1999). Overall superiority of girls has been reported, which is based on girls' drawings being more advanced than boys' in their details or use of lines (Cox, Koyasu, Hiranuma, & Perara, 2001; Mortensen, 1991; Willsdon, 1977). Girls are more likely to combine accurate expression and abstract expression than boys (Picard & Bouhais, 2011). Regarding the spatial structure of drawings, girls tend to place objects in a row or use frontal arrangements whereas the occurrence of midair composition was significantly higher in boys' drawings (Iijima, Arisaka, Minamoto, & Arai, 2001). Girls were good at drawing accurately, whereas boys were more like to preserve the optical appearance of the object array (Lange-Küttner & Ebersbach, 2013). Given the findings of gender differences in creativity (Baer & Kaufman, 2008) and graphic fluency (Picard, 2015), girls have been considered as

having more advantages than boys.

Indeed, this phenomenon is becoming even more problematic if emotional and behavioral difficulties are evaluated using their drawings. Perhaps due to the recognition that boys are generally at higher risk for delinquency than girls (Cohn & Modecki, 2007), it may not be unusual for boys to draw more violent content than girls regardless of their actual condition. With the gradual evolution of projective drawing tests as a standardized diagnostic assessment to screen aggressive or depressive groups, boys have been increasingly considered as a greater risk group than girls because the scoring systems of most projective drawing tests are based on the evaluation of detailed content and structure of lines or composition which are criteria that favor girls. According to a study on depiction of sadness and anger in drawings of the human face, anger is depicted more frequently by boys than girls in response to angry scenarios for both gender characters (Brechet, 2013). If emotions could be evaluated by drawings, boys appear to have more anger than girls.

However, most interpretation of projective drawing tests does not have gender criterion. The rating system of projective drawing assessments is typically based on the examiner's overall impression of a drawing including specific content items and overall organization of the drawings without any criterion considering baseline of gender differences (e.g., Harmon-Walker & Kaiser, 2015). Significant gender

differences were reported in a study to identify children at risk for aggression and depression using the Draw-a-Story (DAS) test (Earwood, Fedorko, Holzman, Montanari, & Silver, 2004), even though the same norms were applied for boys and girls. Although it was reliably reported that the DAS combination scores predicts suicidal risk of middle school students, no statements were mentioned on gender differences of the scores in the results (Park & Kim, 2013). Studies found that the DAS and the Silver Drawing Test (SDT) assess the same constructs supporting the construct (convergent) validity of scores (Silver, 2002). A study of scorer reliability found strong reliability for the emotional content scale as .94 and moderate reliability for the self-image scale as .74 (Silver, 2002). In the case of the Diagnostic Drawing Series (DDS) using the same criteria for both genders to confirm healthy adolescents' artwork, no gender differences were reported despite many more female than males among the participants (Ritnour et al., 2014).

Evidence has revealed in fact, that there are no gender differences in test results of psychological difficulties for children being referred to juvenile court (Herrera & McCloskey, 2001). Brechet (2013) also stated that there are different display rules and gender-emotion stereotypes between boys and girls such as boys should not cry and girls should not get angry; thus, differences are not because boys have more anger or girls are sadder. In a study on girls

with congenital adrenal hyperplasia (Iijima, Arisaka, Minamoto, & Arai, 2001), it was revealed that whatever is shown in drawings is not directly connected to the drawer's gender. The girls with congenital adrenal hyperplasia showed an increase in masculine-typical play behavior in their drawings, and better visual-spatial ability in which boys exhibit stronger performance than girls, even though they have gender identity as girls and are reared as girls. This research showed that the reason gender differences on psychological conditions occur on drawings might not be completely due to psychological conditions, but come from their biological differences, because boys and girls have biologically different feelings and ways of perceiving objects when they draw something.

In another study by Silver (1996), no differences in gender and delinquency were found in the drawings in terms of self-image. However, surprisingly, boys in the non-delinquent group drew proportionally more assaultive relationships than boys in delinquent group. Of this result Silver stated, "Perhaps the finding can be explained by the differences between fantasizing about violence and acting violently. A boy, who has internalized prohibitions against acting out biological drives, may fantasize more than one who commits assaultive acts. It may also be that incarceration for antisocial behavior inhibited expressing assaultive fantasies (p. 549)." In terms of solitary drawing, more than twice as many girls as boys drew sad, isolated, or endangered

solitary subjects regardless of delinquency group. If those scores are valid to evaluate their emotional behavioral difficulties, more boys in the delinquency group should have higher scores on assaultive relationships and solitary subjects. Despite of these gender differences, most projective drawing tests have single criterion in their quantified evaluation system.

Based on this recognition, gender differences in children's drawings may indicate that boys and girls have different feelings, ways of perceiving objects, or types of expression when they draw pictures, not necessarily indicate levels of psychological disturbance. It is therefore important to examine gender differences further and to establish a baseline for projective drawing tests that are frequently used for diagnostic purposes in South Korea. In this study, gender differences in emotional behavioral difficulties evaluated through projective drawing tests were investigated with consideration for their psychological conditions evaluated through self-reported questionnaires. As covariates to check baseline of psychological conditions, automatic positive thought and emotional reactivity were used in this study which are highly correlated with emotional behavioral difficulties. The variables of psychological conditions could work as another assessment to test construct validity. Scores of the Emotional Behavioral Drawing Test (EBDT; Lim, 2014) were used in this study as dependent variables. The EBDT scores are consisted to contents

aspects including aggression and depression, and structure aspects including deficient delineation and imbalance. Emotional behavioral difficulties of boys and girls were reflected to the four sub-factors of the EBDT, supported by projection theory (Lim, 2014).

The purpose of this study was to investigate whether gender differences on the EBDT reflect real gender differences of emotional and behavioral difficulties across boys and girls when gender criterion are applied to scoring. It was hypothesized that the EBDT scores will differ across gender if gender differences are seen with positive automatic thoughts and emotional reactivity across gender. Likewise, it was hypothesized that the EBDT gender differences would be reduced when positive automatic thoughts and emotional reactivity covariates are applied. There were three specific hypotheses in this study, including (a) the EBDT scores will differ significantly between boys and girls based on the scale development study of the EBDT (Lim, 2014); (b) the positive automatic thought and emotional reactivity scores won't be significantly different between boys and girls; (c) there will be no significant effects of positive automatic thought and emotional reactivity covariates on the EBDT scores. Through this study, reliability and validity of the EBDT could be better understood to identify needed scoring criteria score boys versus girls' EBDT drawings.

Method

Participants

A total of 172 children participated in this study. They were recruited from eleven classes of an elementary school in the capital city of South Korea. The participants were from fifth grade with age ranging from eleven to twelve ($N = 62$, 36%) and sixth grade with age ranging from twelve to thirteen ($N = 110$, 64%). Among the participants, 83 (48%) were boys, and 89 (52%) were girls. Nine evaluators participated in scoring 172 drawing tests. All were female, experienced art therapists working for over two years. Six evaluators had 2-3 years working experience, and three evaluators had 3-4 years. All evaluators had training on the EBDT scoring system according to the scale development study of the EBDT (Lim, 2014), and had repeated practice before they gave scores to the participants' drawings. Inter-scorer reliability of each scored item of the EBDT ranged from .52-.97.

Measures

The Emotional Behavioral Drawing Test: EBDT

The Emotional Behavioral Drawing Test (EBDT) was developed by Lim (2014) for Korean children and adolescents. The EBDT employs identical procedures to the Draw a

Story test (DAS; Silver, 1983) using stimulus drawings to elicit responses. Responses of DAS are scored on 5-point rating scales that range from low to high levels of Ability of Select, Combine, and Represent, and from strongly negative to strongly positive Emotional Content, Self-Image, and Appropriate Use of Humor (Silver, 2002). Based on the DAS stimulus cards, thirty stimulus cards were newly developed for Korean children that considered known differences in cultural context. The stimulus cards of EBDT are consisted with human figures (5 cards), animals or natural objects (6 cards), school circumstances (8 cards), and inanimate objects or environments (11 cards). The participants were asked to select 2-5 cards among the 30, and make up stories about those selected cards. The test takes 5-10 minutes to complete.

The EBDT consisted of fifteen scoring items with four factors including hostile expression (which were concerned with aggression and delinquency, named "Aggression" in this study), internal withdrawal (which involved social isolation and depression named "Depression" on this study), structural poverty (which indicated drawings with few and simple details named "Deficiency" in this study), and structural imbalance (which related to overall arrangement of figures and items in drawings named "Imbalance" in this study). The first two factors assess content of drawings represented by dysfunctional attitudes in emotion and behavior,

and later two factors assess structure of drawings that reflected emotional behavioral difficulties.

The scoring system of the EBDT is illustrated in Table 1 with item descriptions of each sub-scale and their reliabilities. Inter-scoring reliabilities for each item in this study were ranged from .52-.97 and are presented in Table 1. It is important to note however that when there is clear scoring on such tests, it is not uncommon for inter-rater reliability scores reaching near perfect agreement. In such situation, it is unclear whether the near perfect agreement is due to the 'true reliability' being near perfect, or due to the scoring rules simply being very clear regardless of the 'true' reliability' of the drawings. For this reason and ambiguity in interpreting high inter-rater reliability results, adequate inter-rater reliability should thus be viewed as a minimum, but not sufficient, bar to pass when testing the psychometric properties of such projective tests. Support for construct validity of the EBDT scores was obtained through exploratory factor analysis ($N = 658$) that identified a four-factor solution and confirmatory factor analysis with all fit indices also indicating adequate fit to the data (Lim & Yang, 2016). According to Lim and Yang (2016), the concurrent validity for the EBDT was demonstrated through its significant correlations with self-or parent-reported emotion and behavioral scales, including the Korean version of Achenbach's Youth Self Report (Oh, Ha, Lee, & Hong, 2001), Korean version of

Rosenberg's Self-Esteem Scale (Lee & Won, 1991), Korean version of Buss and Durkee's Inventory for Assessing Different Kinds of Hostility (Shin & Choi, 2003), and the Bullying-Behavior Scale (Lee & Kwak, 1999).

Construct validity was established through confirmatory factor analysis and concurrent validity of the EBDT was found to be adequate (Lim, 2014). All sub-scales of the EBDT had significant correlations with Withdrawal, Somatization, and Anxiety / Depression of the Youth Self-Report (Achenbach, 1991), three coping strategies of the Brief COPE (Carver, 1997), five types of aggression from the Inventory for Assessing Different Kinds of Hostility (Buss & Durkee, 1957), and Assault and Damage of the Peer Victimization Behavioral Scale (Callaghan & Joseph, 1995). The EBDT was used as a quantified projective drawing test to measure emotional and behavioral difficulties in this study. Since the EBDT is a recently validated instrument only in the Korean population, its limitation was addressed in the discussion section.

The Automatic Thought Questionnaire-Positive: ATQ-P

Developed by Ingram and Wisnicki (1988) to assess the frequency of positive self-statements or thoughts repeatedly, the Automatic Thought Questionnaire-Positive (ATQ-P) was conducted to assess the participants' positive automatic cognition. As a central construct in

psychopathology, automatic thinking is repetitive, intrusive, and relatively uncontrollable self-relevant thoughts. Deficits in positive cognition may automatic thought. The participants were asked how frequently they thought about the listed items during the last week and asked to respond with a 5-point Likert scale. The ATQ-P includes three sub-factors such as positive daily functioning (e.g., “I’m comfortable with life”, “I take good care of myself”), evaluations of self (e.g., “I have good friends who support me”, “I have a good way with others”), and positive

future expectation (e.g., “My future looks bright”, “My life keeps getting better”). The internal consistencies for each sub-factor in this study were reported as .90, .88, and .92.

The ATQ-P for youth has significant negative correlations with the Children’s Depression Inventory (Kovacs, 1981), the Revised Children’s Manifest Anxiety Scale (Reynolds & Richmond, 1978), the Negative Affect Self-Statement Questionnaire (Ronan, Kendall, & Rowe, 1994), and the Children’s Automatic Thoughts Scale (Schniering & Rapee, 2002). The ATQ-P was

Table 1. Subscales and Reliabilities of the EBDT

Sub-factors	N	Item Descriptions (Scoring from 0 to 3)	Inter-rater <i>a</i>	<i>a</i>		
CONTENTS	AGG.	4	Aggressive emotions	.97	.86	
			Harassing social relationship between characters	.94		
			Assaulted expression including murder, delinquency, and violation	.95		
	DEP.	4	Anger and negative emotions	.91	.63	
				Lack of social ability, passive or unilateral relationship, withdrawal		.76
				Sadness, loneliness, or sinking feelings		.72
STRUCTURE	DEF.	3	Sense of inferiority, self-devaluation, self-reproach, or guilty	.52	.67	
				Tiredness, helplessness, or boredom		.71
	IMB.	4	Less than 5 cards, additional descriptions or decorations	.84	.62	
				Number of objects		.79
				Replace cards with writings or drawings, less than 2 cards, no additions		.85
		Use space less than 25%	.82			
		Location of center beyond 6cm of radius	.81			
		Inappropriate interaction between cards	.64			
		Unbalanced composition and proportion	.65			
Total				.64		

measured in order to control the initial difference due to the effects of covariates. The Korean version of the ATQ-P validated by Lee and Kim (2002) was used in this study. The reliability and validity of the Korean version of the ATQ-P for youth group were confirmed with The Cronbach's α was .97. The internal consistencies for each sub-factor of the Korean version were .87 (positive daily functioning), .84 (evaluations of others), and .93 (positive future expectation; Yang, Hong, Jung, & Kim, 2005).

Emotional Reactivity: Differentiation of Self Inventory-Revised (DSI-R)

In the Differentiation of Self Inventory-Revised (DSI-R), developed by Skowron and Schmitt (2003), 11 items from the subscale "Emotional Reactivity" were used in this study because these 11 items were designed to measure the intra-psychoic dimension of self which is the ability to regulate emotion (Skowron & Schmitt, 2003). The Emotional Reactivity subscale included the items to assess the response towards environmental stimuli on the basis of autonomic emotional responses with items such as "If someone is upset with me, I can't seem to let it go easily" or "I react impulsively to situations and regret my actions later." The participants rated items using a 6-point Likert scale, ranging from 1 (*not at all true of me*) to 6 (*very true of me*).

The Emotional Reactivity of DSI-R was used as a covariance factor in this study since it has

had significant correlations with perceived stress and psychological distress (Krycak, Murdock, & Marszalek, 2012). The Korean version of the DSI-R for 4th and 5th graders validated by Koo and Kim (2014) was used in this study. The internal consistence of the Korean version of the DSI-R for children was reported as .80. The Cronbach's α of the 11-items from the sample of Korean children in this study was .82. The 11-items from the DSI-R were measured for the same reason with the ATQ-P in order to control the initial difference due to the effects of covariates.

Data Analysis

To observe the baseline of gender differences between boys and girls, independent t-tests were conducted for every measured variable. To examine the gender differences in the Emotional Behavioral Drawing Test (EBDT), MANCOVA (multiple analyses of covariance) procedures were employed to guard against type I error that might occur if multiple ANCOVAs were conducted independently. MANCOVA is an extension of ANCOVA with several dependent variables. MANCOVA is useful because it can reveal differences not discovered by ANCOVA tests. In addition, because the four dependent variables, the subscales of EBDT (Aggression, Depression, Deficiency, and Imbalance) in this study were conceptually related to each other as suggested by the correlation coefficients (.187 to

.497), the MANCOVA procedure was more suitable for this type of analysis that controlled correlations among dependent variables. The within-subject variable, two covariates, Positive Automatic Thought and Emotional Reactivity, served as control variables to reduce the initial difference due to the effects of covariates in the model. The independent variable (i.e., child's gender) was coded at two levels: Level 1 for boy and Level 2 for girl.

Results

As shown in Table 2, independent t-tests confirmed there were no statistically significant differences between boys and girls on their psychological conditions such as positive

automatic thoughts and emotional reactivity. However, significant differences of the EBDT scores between boys and girls were revealed, except for Deficiency. Hereby the two hypotheses were supported. The Table 3 showed that the third hypothesis was supported. Table 3 presents the correlations among ATQ-P, Emotional Reactivity of DSI-R, and EBDT. The results showed that the relationship between the EBDT scores and self-reported psychological conditions are not significant. As shown in Table 4, MANCOVA results indicated statistically significant gender differences in EBDT scores [Wilks' Lambda = .919, $F(4, 165) = 3.61, p < .01$]. However, there were no significant effects of covariates, both Positive Automatic Thoughts [Wilks' Lambda = .973, $F(4, 165) = 1.16, p = .33$] and Emotional Reactivity

Table 2. Mean Differences of Measured Variables between Boys and Girls

	Boys (N = 83)		Girls (N = 89)		t
	M	SD	M	SD	
ATQ_P-Total	3.65	.83	3.78	.67	-1.14
Daily Functioning	3.44	.91	3.50	.77	-.50
Evaluations of Self	3.70	.80	3.91	.70	-1.85
Future Expectation	3.76	.97	3.90	.75	-1.02
Emotional Reactivity of DSI-R	3.26	.43	3.33	.43	-1.06
EBDT-Total	46.57	19.38	35.69	16.57	3.95***
Aggression	8.15	12.10	3.47	7.40	3.14**
Depression	13.48	7.12	10.96	7.31	2.64**
Deficient Delineation	13.40	4.09	12.82	4.53	1.23
Imbalance	11.18	5.86	8.79	5.37	2.81**

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

Table 3. Correlations among sub-scales of the ATQ-P, ER of the DSI-R, and sub-scales of the EBDT for Boys and Girls

	1	2	3	4	5	6	7	8
1. Aggression (EBDT)	1	.52**	-.15	.13	.01	-.20	-.22*	-.01
2. Depression (EBDT)	.44**	1	-.01	.32**	-.02	-.20	-.18	-.13
3. Deficiency (EBDT)	-.04	.15	1	.17	-.04	.04	.01	-.05
4. Imbalance (EBDT)	.19	.42**	.31**	1	.14	.09	.09	.11
5. Daily Functioning (ATQ-P)	-.10	-.17	-.11	.03	1	.70**	.73**	.55**
6. Evaluations of Self (ATQ-P)	.10	.13	-.10	.10	.55**	1	.76**	.60**
7. Future Expectations (ATQ-P)	-.12	-.13	-.06	-.03	.74**	.75**	1	.54**
8. Emotional Reactivity (DSI-R)	-.04	-.06	-.11	-.02	.59**	.41**	.55**	1

* $p < .05$, ** $p < .01$, N: boys = 83, girls = 89

Note. Inter-correlations among boys are above the diagonal: inter-correlations among girls are below the diagonal.

Table 4. Means (standard error) for dependent variables by gender

Dependent Variables	Gender ^a		MANCOVA			
	Boys (N = 83)	Girls (N = 89)	F	η_p^2 ^b	Observed Power ^c	Cohen's d ^d
Aggression	8.15 (12.10)	3.47 (7.40)	9.170**	.052	.853	.47
Depression	13.48 (7.12)	10.96 (7.31)	5.087*	.029	.61170	.35
Deficiency	13.40 (4.09)	12.82 (4.53)	.745	.004	.138	.13
Imbalance	11.18 (5.86)	8.79 (5.37)	7.410**	.042	.772	.43

* $p < .05$, ** $p < .01$

Note.

^a Adjusted mean values after controlling for Positive Automatic Cognition and Differentiation of Self

^b Partial eta squared

^c Computed using alpha = .05

^d Cohen's d based on adjusted means and pooled standard deviations.

{Wilks' Lambda = .975, $F(4, 165) = 1.07$, $p = .37$ }.

Further analyses (univariate analyses to evaluate one variable at a time) indicated that the independent variable (gender) was a significant factor on three dependent variables, the three subscales of EBDT (Aggression, Depression, and Imbalance). The Least Significant Difference method (LSD; post hoc-method that can handle both pairwise and non-pairwise comparisons and does not require equal sample sizes) was used to examine mean differences for boys and girls. Post-hoc comparisons were used to explore differences, not limited by ones specified in advance on the basis of theory. As shown in Table 2, the post-hoc analysis indicates that boys have higher scores than girls on three dependent variables (Aggression, Depression, and Imbalance). On the other hand, no significant gender difference was found on Deficiency, one of the EBDT subscales.

In addition to significant tests for differences (i.e., LSD), the Cohen's d (Cohen, 1988) was used to calculate the effect size (strength of differences) between boys and girls. Effect sizes were calculated to compensate for the large sample size. Cohen (1988) defined effect sizes as "small, $d = .2$," "medium, $d = .5$," and "large, $d = .8$ ". As shown in Table 2, for example, an effect size .47 (medium effect) indicates that the boys ($M = 8.15$, $SD = 12.10$) have higher scores than the girls ($M = 3.47$, $SD = 7.40$) on Aggression. On the

Depression scale, the effect size .35 (small to medium effect) indicates that the boys ($M = 13.48$, $SD = 7.12$) have higher scores than the girls ($M = 10.96$, $SD = 7.31$). Lastly, on the Imbalance scale, the effect size .43 (medium effect) indicates that the boys ($M = 11.18$, $SD = 5.86$) have higher scores than the girls ($M = 8.79$, $SD = 5.37$) on Imbalance.

With a Cohen's d of .50, approximately 70% of the boys would be above the mean of the girls in the EBDT. Therefore, there is a 70% chance that a person picked at random from the boys would have a higher score in EBDT than a person picked at random from the girls (probability of superiority). Overall, the results of the present study indicate that boys are more likely to be evaluated as more psychologically problematic than girls as measured by the EBDT with no covariate effect of positive psychological conditions such as positive automatic thoughts and emotional reactivity.

Discussion

This study was conducted to test the gender effects on the projective drawing test named EBDT, used to evaluate emotional behavioral difficulties. In order to determine whether higher scores from the EBDT for boys than girls actually reflect higher risk of psychological problems, positive automatic thoughts and emotional reactivity were measured to control for

the initial difference as effects of covariates. Although gender differences of psychological condition on projective drawing tests were repeatedly reported, there has been little literature to test whether the differences between boys and girls on projective drawing tests are based on actual differences of their psychological condition. Based on this recognition, this study was conducted.

The results indicated statistically significant gender differences in EBDT scores. Specifically, in the research findings, boys had significantly higher scores than girls in terms of emotional behavioral problems. Despite showing adequate functioning in positive automatic thought and emotional reactivity, scores of the EBDT for boys indicated that boys are more problematic than girls emotionally and behaviorally. Interestingly, there were no statistically significant differences between boys and girls on psychological conditions such as positive automatic thoughts and emotional reactivity. Also, significant differences of the EBDT scores between boys and girls were revealed except for deficient delineation. While there were no significant effects of covariates, Positive Automatic Thoughts and Emotional Reactivity, gender was a significant factor on three dependent variables, Aggression, Depression, and Imbalance. Boys have higher scores on aggression, depression and imbalance. No significant gender difference was found on Deficiency, one of the EBDT subscales.

Given the results of the study, it is possible that boys who draw more aggressive or depressive drawings than girls could actually be no different from girls on psychological conditions. It is repeatedly recommended that there is need to apply gender criterion to score all types of drawing tests because boys and girls are different in drawings and expressing their emotions regardless of their psychological conditions. Extant literature offers various ways to address this persistent problem.

Although Cohen, Hammer, and Singer (1988) notably asserted that intuitive process is an important factor in psychotherapy particularly in their interpretive work, practitioners must engage in their efforts to accurately evaluate and diagnose children for psychological conditions in the course of using projective drawing tools, with particular care in their assessment of boys given the critiques put forth in the literature and the lack of gender criterion of projective drawing tests. Additionally, Betts (2006) argued that the best approach to assessment includes objective testing and elicitation of the subjective viewpoints of the client. The value of a client's subjective appraisal of their own drawing should not be underestimated. Betts contends that "... those who choose to assess clients through art have neglected to convincingly address the essence of empirical scientific inquiry-findings that link character traits with artistic expressions; replicable results based upon copious and random data; and, uniform outcome measures

that justify diagnosis of a client via his or her artwork (p.427).”

There are some limitations in this study that warrant mentioning. Positive automatic thoughts and emotional reactivity were measured to control for the initial differences due to the effects of covariates. In retrospect, aggression and depression should be considered measuring as covariant variables to learn more about the effect of gender on aggression and depression of the EBDT. Indeed, the correlations between the EBDT scores and psychological conditions such as automatic positive thought and emotional reactivity were not significant. It could make the construct validity of the EBDT worse. However, before to consider risk of validity, it should have made more careful choice for the ATQ-P and DSI-R. It is major limitation of this study.

Additionally, relatively new projective test was used only validated in a Korean children’s sample. Further analyses should be conducted with scales validated across cultures since scholars have reported nuances in responses by children from multiple nations (e.g., Alter-Muri & Vazzano, 2014). Finally, other projective tests should be evaluated with the aim to establish much needed gender norms within and across cultures. The present study revealed that despite serious weaknesses on the reliability and validity of their scoring system (Lilienfeld, Wood, & Garb, 2000); projective tests yield higher problematic scores for boys. These limitations inform recommendations for future research.

Based on extant literature and results from this study, the need remains to conduct further research that would help establish gender norms for quantification of scores in projective drawing tests such as the EBDT. Neale and Rosal (1993) outlined robust research methods that should be implemented when conducting research with projective tests and they remain valuable today.

Given the relatively weak evidence for the incremental validity of most projective indexes, the Clinical Psychology Division of American Psychological Association long ago recommended to exclude training in projective techniques from the graduate assessment curriculum (Grove, 2000), or to make students aware of the negligible relationship between the amount of prior experience with an assessment technique and its predictive accuracy (Garb, 1998) if instructors intend to cover them. Although criticisms and recommendations are very strong against their use, it is generally believed that there is considerable relationship between peculiarity in drawings and psychopathological characteristics, and the incidents may be more powerful in real world situations than in experimental settings (Kubiszyn et al., 2000). For these reasons, Lilienfeld (1999) suggested that the long and strict training for clinicians who use projective drawing tests necessitates mastering “a skill that does not come naturally to any of us: disregarding the vivid and compelling data of subjective experience in favor of the often dry and impersonal results of

objective research” (p. 38). His statement highlights a significant implication for psychological counselors who are the most frequent users of projective drawing tests. Finally, results from this research underscore that although projective drawing tests remain useful and may be applied effectively, practitioners need to be especially aware of gender differences on artistic expression and potential for labeling boys negatively based on typical drawing practices for their gender.

References

- Achenbach, T. M. (1991). *Integrative Guide for the 1991 CBCL/4-18, YSR, and TRF Profiles*. Burlington: Department of Psychiatry, University of Vermont.
- Alter-Muri, S. B., & Vazzano, S. (2014). Gender typicality in children’s art development: A cross-cultural study. *The Arts in Psychotherapy, 41*(2), 155-162.
- Arteche, A., & Murray, L. (2011). Maternal affective disorder and children’s representation of their families. *Journal of Child and Family Studies, 20*(6), 822-832.
- Baer, J., & Kaufman, J. C. (2008). Gender differences in creativity. *The Journal of Creative Behavior, 42*(2), 75-105.
- Betts, D. J. (2006). Art therapy assessments and rating instruments: Do they measure up? *The Arts in Psychotherapy, 33*(5), 422-434.
- Blakemore, J. E. O., Berenbaum, S. A., & Liben, L. S. (2009). *Gender Development*. New York: Taylor & Francis.
- Boyatzis, C. J., & Eades, J. (1999). Gender differences in preschoolers’ and kindergartners’ artistic production and preference. *Sex Roles, 41*(7-8), 627-638.
- Brechet, C. (2013). Children’s gender stereotypes through drawings of emotional faces: Do boys draw angrier faces than girls? *Sex Roles, 68*(5-6), 378-389.
- Burns, R. C. (1987). *Kinetic House-Tree-Person (KHTP) drawings: An interpretative manual*. New York: Brunner.
- Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology, 21*(4), 343-349.
- Callaghan, S., & Joseph, S. (1995). Self-concept and peer victimization among schoolchildren. *Personality and Individual Differences, 18*(1), 161-163.
- Carver, C. C. (1997). You want to measure coping but your protocol’ too long: Consider the brief cope. *International Journal of Behavioral Medicine, 4*(1), 92-100.
- Cherney, I. D., & London, K. (2006). Gender-linked differences in the toys, television shows, computer games, and outdoor activities of 5-to 13-year-old children. *Sex Roles, 54*(9-10), 717-726.
- Cherney, I. D., Seiwert, C. S., Dickey, T. M., & Flichtbeil, J. D. (2006). Children’s drawings: A mirror to their minds. *Educational Psychology, 26*(1), 127-142.
- Cohen, B. M., Hammer, J. S., & Singer, S. (1988). The diagnostic drawing series: A

- systematic approach to art therapy evaluation and research. *The Arts in Psychotherapy*, 15(1), 11-21.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohn, E. S., & Modecki, K. L. (2007). Gender differences in predicting delinquent behavior: Do individual differences matter? *Social Behavior and Personality*, 35(3), 359-374.
- Cox, M. V., Koyasu, M., Hiranuma, H., & Perara, J. (2001). Children's human figure drawings in the UK and Japan: The effects of age, sex, and culture. *Developmental Psychology*, 19(2), 275-292.
- Drake, M., Lo, J. L., Hwang, M. T., & Shih, L. (1995). Comparison of color meaning and content of drawings by Taiwanese & USA psychiatric patients. *Journal of Taiwan Occupational Therapy Association*, 12, 41-49.
- Earwood, C., Fedorko, M., Holzman, E., Montanari, L., & Silver, R. (2004). Screening for aggression using the draw a story assessment. *Art Therapy*, 21(3), 156-162.
- Garb, H. N. (1998). Recommendations for training in the use of the Thematic Apperception Test (TAT). *Professional Psychology: Research and Practice*, 29(6), 621-622.
- Goodenough, F. L. (1926). *Measurement of Intelligence by Drawings*. Oxford: World Book Co.
- Groth-Marnat, G., & Roberts, L. (1998). Human figure drawings and House Tree Person (KHTP) drawings as indicators of self-esteem: a quantitative approach. *Journal of Clinical Psychology*, 54(2), 219-222.
- Grove, W. M. (2000). *APA Division 12 Presidential Task Force "Assessment for the Year 2000 Report of the Task Force"*. Washington, D. C.: American Psychological Association, Division 12 (Clinical Psychology).
- Halpern, D. F. (2004). A cognitive-process taxonomy for sex differences in cognitive abilities. *Current Directions in Psychological Science*, 13(4), 135-139.
- Harmon-Walker, G. & Kaiser, D. H. (2015). The bird's nest drawing: A study of construct validity and interrater reliability. *The Arts in Psychotherapy*, 42, 1-9.
- Harris, D. B. (1963). *Drawing as a measure of intellectual maturity: A revision and extension of the Goodenough Draw-a-Man test*. New York: Harcourt, Brace & World.
- Herrera, V. M., & McCloskey, L. A. (2001). Gender differences in the risk for delinquency among youth exposed to family violence. *Child Abuse & Neglect*, 25(8), 1037-1051.
- Howard, A. R. H., Razuri, E. B., Call, C. D., DeLuna, J. H., Purvis, K. B., & Cross, D. R. (2017). Family drawings as attachment representations in a sample of post-institutionalized adopted children. *The Arts in Psychotherapy*, 52, 63-71.
- Iijima, M., Arisaka, O., Minamoto, F., & Arai, Y. (2001). Sex differences in children's free drawings: A study on girls with congenital adrenal hyperplasia. *Hormones and Behavior*, 40(2), 99-104.
- Ingram, R. E., & Wisnicki, K. S. (1988). Assessment of positive cognition. *Journal of Consulting and Clinical Psychology*, 56(6),

- 898-902.
- Knoff, H. M., & Prout, H. T. (1985). *Kinetic drawing system for family and school: A handbook*. Los Angeles, CA: Western Psychological Services.
- Koo, K., & Kim, S. (2014). Effects of family strength on subjective well-being perceived by elementary students: Mediating effect of self-differentiation and moderating effect of gender difference. *The Korean Journal of Youth Research, 21*(6), 281-307.
- Koppitz, E. M. (1968). *Psychological evaluation of children's human figure drawings*. London: Grune & Stratton.
- Kovacs, M. (1981). Rating scales to assess depression in school-aged children. *Acta Paedopsychiatrica, 46*(5-6), 305-315.
- Krycak, R. C., Murdock, N. L., & Marszalek, J. M. (2012). Differentiation of self, stress, and emotional support as predictors of psychological distress. *Contemporary Family Therapy, 34*(4), 495-515.
- Kubiszyn, T. W., Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., & Eisman, E. J. (2000). Empirical support for psychological assessment in clinical health care settings. *Professional Psychology: Research and Practice, 31*(2), 119-130.
- Lange-Küttner, C., & Ebersbach, M. (2013). Girls in detail, boys in shape: Gender differences when drawing cubes in depth. *British Journal of Psychology, 104*(3), 413-437.
- Lee, C. J., & Kwak, K. J. (1999). *Bullying in school: Conditions and characteristics*. Seoul: Jipmoondang.
- Lee, H. J., & Won, H. T. (1991). An investigation of the relationship among paranoid tendency, self-concept, and self-awareness. *The Korean Journal of Psychological Association, 6*, 277-290.
- Lee, J. Y., & Kim, J. H. (2002). Assessment of positive cognition: A standardization study of Korean Automatic Thought Questionnaire-Positive: ATQ-P. *Korean Journal of Clinical Psychology, 21*(3), 647-664.
- Lilienfeld, S. O. (1999). Projective measures of personality and psychopathology: How well do they work? *Skeptical Inquirer, 23*, 32-39.
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Health, 1*(2), 27-66.
- Lim, N. Y., & Yang, E. (2016). Development and validation of drawing test of youth emotional and behavior problems (Emotional & Behavioral Drawing Test: EBDT). *The Korean Journal of Arts Therapy, 16*(1), 1-16.
- Lim, N. Y. (2014). *Development and validation of drawing test of youth emotional behavioral problems*. Unpublished Doctoral Dissertation, Korean University, Seoul, Korea.
- Malchiodi, C. (1990). *Breaking the silence: Art therapy with children from violent homes*. New York: Brunner/Mazel.
- Marzolf, S. S., & Kirchner, J. H. (1973). Personality traits and color choices for House-Tree-Person drawings. *Journal of Clinical Psychology, 29*(2), 240-245.
- Matlin, M. W. (2004). *The psychology of women* (5th ed.). Belmont, CA: Thomson/Wadsworth

- Publishers.
- Mortensen, K. V. (1991). *Form and content in children's human figure drawings: Development, sex differences, and body experience*. New York, NY: New York University Press.
- Naglieri, J. A. (1988). *DAP: Draw A Person, a quantitative scoring system manual*. New York: The Psychological Corporation.
- Neale, E. L., & Rosal, M. L. (1993). What can art therapists learn from the research on projective drawing techniques for children? A review of the literature. *The Arts in Psychotherapy*, 20(1), 37-49.
- Oh, K. J., Ha, E. H., Lee, H. R., & Hong, K. (2001). *K-YSR Youth Self Report*. Seoul: Chung-ang Aptitude Publisher.
- Palmer, L., Farrar, A. R., Valle, M., Ghahart, N., Panella, M., & DeGraw, D. (2000). An investigation of the clinical use of the House-Tree-Person projective drawings in the psychological evaluation of child sexual abuse. *Child Maltreatment*, 5(2), 169-175.
- Park, C. H., & Kim, G. S. (2013). A validation study on DAS in the prediction of suicidal risk for adolescents. *The Arts in Psychotherapy*, 40(1), 108-114.
- Pianta, R. C., Longmaid, K., & Ferguson, J. E. (1999). Attachment-based classifications of children's family drawings: Psychometric properties and relations with children's adjustment in kindergarten. *Journal of Clinical Child Psychology*, 28(2), 244-255.
- Picard, D. (2015). Sex differences in scores on the Draw-A-Person test across childhood: Do they relate to graphic fluency? *Perceptual and Motor Skills*, 120(1), 273-287.
- Picard, D., & Bouhais, M. (2011). Sex differences in expressive drawing. *Personality and Individual Differences*, 51(7), 850-855.
- Reynolds, C. R., & Richmond, B. O. (1978). What I think and feel - a revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology*, 6(2), 271-280.
- Ritnour, M. M., Bovingdon, M., Knutsen, C., Roy, A., Hoshino, J., & Johnson, K. (2014). An examination of differences between adults and adolescent controls using the Diagnostic Drawing Series. *The Arts in Psychotherapy*, 42, 63-74.
- Ronan, K. R., Kendall, P. C., & Rowe, M. (1994). Negative affectivity in children: Development of validation of a self-statement questionnaire. *Cognitive Therapy and Research*, 18(6), 509-528.
- Rothkopf, I., & Turgeon, S. M. (2014). Sex differences in college students' free drawings and their relationship to 2D:4D ratio and recalled childhood play behavior. *Archives of Sexual Behavior*, 43(2), 311-318.
- Schniering, C. A., & Rapee, R. M. (2002). Developing and validity of a measure of children's automatic thoughts: The children's automatic thoughts scale. *Behavior Research and Therapy*, 40, 1091-1190.
- Shin, H. Y., & Choi, H. R. (2003). The relations among child abuse by parents, child's aggression and interpersonal anxiety. *The Korean Journal of Counseling and Psychotherapy*, 15, 295-307.
- Silver, R. (1996). Sex differences in the solitary and assaultive fantasies of delinquent and

- non-delinquent adolescents. *Adolescents*, 31(123), 543-553.
- Silver, R. A. (1983). *Silver drawing test of cognition and emotion*. Seattle, WA: Special Child Publications.
- Silver, R. A. (1993). *Draw-A-Story: Screening for depression and age or gender differences*. Sarasota, FL: Ablin Press Distributors.
- Silver, R. A. (2002). *Three art assessments: The Silver drawing test of cognition and emotion; Draw A Story, screening for depression; and stimulus drawings & techniques*. New York: Brunner-Routledge.
- Skowron, E. A., & Schmitt, T. A. (2003). Assessing interpersonal fusion: Reliability and validity of a new DSI fusion with others subscale. *Journal of Marital and Family Therapy*, 29(2), 209-222.
- Turgeon, S. M. (2008). Sex differences in children's face drawings and their relationship to 2D:4D ratio. *Personality and Individual Differences*, 45, 527-532.
- Willsdon, J. A. (1977). *Discussion of same sex differences in a study of human figure drawings by children aged four-and-a-half to seven-and-a-half years*. In G. Butterworth (Ed.), *The child's representation of the world*. New York: Plenum.
- Yang, J. W., Hong, S. D., Jung, Y. S., & Kim, J. H. (2005). Validation study of automatic thoughts questionnaire (ATQ-N) and the positive automatic thoughts questionnaire (ATQ-P) in adolescents. *The Korean Journal of Clinical Psychology*, 24(3), 631-646.
- Zalsman, G., Netanel, R., Fischel, T., Freudenstein, O., Landau, E., Orbach, I., et al. (2000). Human figure drawings in the evaluation of severe adolescent suicidal behavior. *American Academy of Child and Adolescent Psychiatry*, 39(8), 1024-1031.
- 원 고 접 수 일 : 2017. 08. 16
수정원고접수일 : 2017. 10. 25
게 재 결 정 일 : 2017. 11. 21

초등학생들의 정서행동문제 평가를 위한 투사적 그림검사에서 나타나는 성별효과의 오류

안 성 희

홍익대학교

상담심리학분야의 실무자들이 흔히 사용하는 투사적 그림검사는 객관적 평가의 준거가 부족함에도 불구하고 내담자의 심리적 상태를 알아보기 위해 빈번하게 사용되어왔다. 그림검사를 토대로 진행된 많은 선행연구에서 남아들이 여아들보다 정서적·행동적으로 더 많은 문제를 가지고 있음이 보고되었으나 이는 성별의 차이를 고려하지 않은 채점 준거에 근거한 결과인 경우가 많았다. 이 연구에서는 투사적 그림검사에서 나타나는 정서행동문제의 성별의 차이가 과연 실제 존재하는 성별의 차이에 의한 것인지를 알아보았다. 서울시내 초등학교 5, 6학년에 재학 중인 172명의 남녀 아동들을 대상으로 정서행동그림검사(EBDT)를 사용하여 채점자 6인의 점수를 토대로 아동들의 정서행동상의 문제를 점수화하였다. 더불어 긍정적 자동적사고와 정서적 반응성을 측정하여 아동들의 심리적 상태의 기저선으로 확인하였다. 그림검사에 나타난 남녀아동들의 정서행동점수는 남녀간 유의미한 차이가 있었으며 긍정적 자동적 사고와 정서적 반응성의 공변량 효과는 유의미하지 않았다. 이 결과는 남아들이 여아들보다 그림검사 상에서 정서행동상의 문제가 더 많은 것으로 나타나더라도 실제로는 두 성별 간 차이가 없을 수 있다는 점을 시사한다. 남녀아동들의 그림발달단계가 다르고 정서적 표현 또한 차이가 있기 때문에, 본 연구의 결과를 토대로 투사적 그림검사에서 남녀별 채점 준거를 마련할 필요성에 대해 피력하였다.

주요어 : 성차, 투사적 그림검사, 정서행동문제