

# The Wartegg Zeichen Test: A Literature Overview and a Meta-Analysis of Reliability and Validity

Jarna Soilevuo Grønnerød  
Fredrikstad, Norway

Cato Grønnerød  
University of Oslo

All available studies on the Wartegg Zeichen Test (WZT; Wartegg, 1939) were collected and evaluated through a literature overview and a meta-analysis. The literature overview shows that the history of the WZT reflects the geographical and language-based processes of marginalization where relatively isolated traditions have lived and vanished in different parts of the world. The meta-analytic review indicates a high average interscorer reliability of  $r_w = .74$  and high validity effect size for studies with clear hypotheses of  $r_w = .33$ . Although the results were strong, we conclude that the WZT research has not been able to establish cumulative knowledge of the method because of the isolation of research traditions.

**Keywords:** Wartegg, meta-analysis, reliability, validity, history

The Wartegg Zeichen Test (WZT, or Wartegg Drawing Completion Test) was introduced by Ehrig Wartegg (1939) as a method of personality evaluation within the Gestalt psychological tradition in Leipzig, Germany (on the early history of the method, see Klemperer, 2000; Lockot, 2000; Roivainen, 2009). The WZT form consists of a standard A4-sized paper sheet with eight 4 cm × 4 cm squares in two rows on the upper half of the sheet. A simple sign is printed in each of the squares (see Figure 1). The test person's task is to make a complete drawing using the printed sign as a part of the picture (see Figures 2 and 3) and then give a short written explanation or title of each drawing on the lower part of the sheet.

Wartegg's (1939) early work includes of a presentation of how different personality types (synthesizing, analytical, and integrated) react in different ways to the small and simple geometrical figures, producing drawings according to the person's typical ways of perceiving and reacting (see also Roivainen, 2009; Wass & Mattlar, 2000). Theoretically, the Wartegg traditions can be categorized into analytical systems of interpretation, which regard the printed signs as visual stimuli (e.g., Kukkonen, 1962a; Takala, 1957; Takala & Hakkarainen, 1953), and dynamic systems (e.g., Crisi, 1998; Gardziella, 1985; Kinget, 1952; Lossen & Schott, 1952; Wass & Mattlar, 2000), which argue that the printed signs have certain symbolic meanings representing certain areas of individual psychology (Tamminen & Lindeman, 2000). As an example of the latter, Kinget (1952) proposed that the printed signs in Square 3 give an impression of rigidity, order, and progression, linking the interpretation of the test persons' drawings to achievement motivation. Similarly, the dot in Square 1 is placed right in the middle, linking the interpretation to images of the self. The symbolic hypothesis is problematic, however, and has been criticized for the lack of empirical verification (Tamminen & Linde-

man, 2000). In general, the theoretical groundwork of the Wartegg method is inadequate, reflecting its development within scattered traditions. We explain this historical process in more detail later in this article.

The methods of interpreting the WZT protocols vary from approaches emphasizing qualitative interpretation (e.g., Avé-Lallemant, 1978; Gardziella, 1985; Wartegg, 1953) to more quantitative scoring systems (e.g., Crisi, 1998, 1999, 2008; Kinget, 1952; Puonti, 2005; Takala, 1957; Wass & Mattlar, 2000). The scoring categories of drawing performance include drawing time, the order of the squares drawn, possible refusals, the size of the drawings, the content of the drawings, crossing of the borders of the squares, shading, drawing line quality, and the written title of the drawing. Although many elements are common, slight differences in scoring definitions between authors and traditions occur. The interpretation is related to many different, but unfortunately not always commonly used or well defined, personality functions. For example, in Wass and Mattlar's (2000) scoring system, which is based on Gardziella (1985), the personality characteristic variables are the following: vitality, initiation and activity, ambition, expansion, spontaneity, energy, ego strength, ego control, independency, objectivity, subjectivity, interest in emotional interaction, emphatic ability, and egoism. The Crisi (2008) system produces three types of evaluations of an individual. First, a qualitative description on eight personality areas corresponding with each WZT square; second, a three-level classification of the individual's emotional, cognitive, and social maturity (reached, partially reached, not reached); and finally, a clinical evaluation (well-structured personality, need for a closer evaluation, and psychopathological condition).

The appropriateness of the drawing in relation to the printed sign is a key aspect in many scoring systems, the basic rationale being that the ability to perceive and respond to the test stimuli corresponds to behavior in social environments. The WZT thus falls into the category of a "projective" test, or what now more precisely should be labeled a performance-based (Meyer & Kurtz, 2006) or free-response method. It seems to have the same type of attraction as the Rorschach method in that the ideographical nature

---

This article was published Online First November 7, 2011.

Jarna Soilevuo Grønnerød, Fredrikstad, Norway; Cato Grønnerød, Department of Psychology, University of Oslo, Oslo, Norway.

Correspondence concerning this article should be addressed to Cato Grønnerød, Department of Psychology, University of Oslo, P.O. Box 1094 Blindern, NO-0317 Oslo, Norway. E-mail: jarna.soilevuo@gronnerod.net

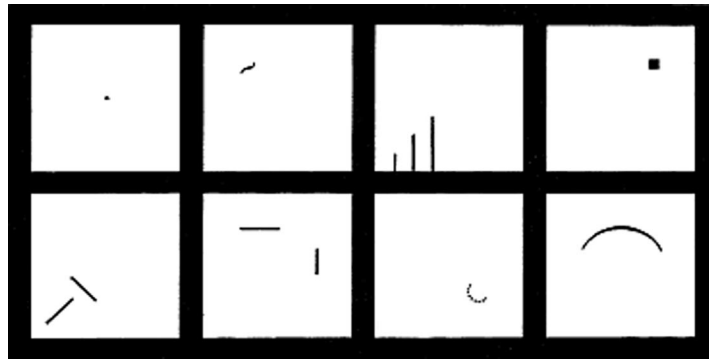


Figure 1. The Wartegg Zeichen Test form. Copyright 1997 Hogrefe Verlag GmbH & Co. KG, Göttingen, Germany. Reprinted with permission.

of the material appeals to the interpretative skills of clinicians. European in origin, the Rorschach method was first introduced in the United States by Beck (1930) and Hertz (1935). The method has since evolved through the interplay of theory and empirical work expressed in the major Rorschach scoring systems. Although many challenges still lie ahead (Meyer & Archer, 2001), developments in the last decade have established the Rorschach method as a personality assessment instrument with empirical backing similar to that of other instruments (Society for Personality Assessment, 2005).

The scientific development of the WZT is markedly different, however, from the Rorschach method. In this article, we argue that the Wartegg traditions in different non-English-speaking countries have evolved in relative isolation from each other, with only occasional contact, and that empirical research has not managed to create cumulative knowledge on the reliability and validity of the method variables. Also, the theoretical and methodological quality of empirical Wartegg studies varies considerably, which is obviously highly problematic for a method used in psychological practice.

Recent surveys among practicing psychologists have confirmed that the Wartegg method is commonly used in Brazil (de Godoy & Noronha, 2005; de Oliveira, Noronha, Dantas, & Santarem, 2005; Noronha, Primi, & Alchieri, 2005; Pereira, Primi, & Cobêro, 2003), Finland (Kuuskorpi & Keskinen, 2008), Italy (Ceccarelli,

1999, as cited in Roivainen, 2009), and Switzerland (Deinlein & Boss, 2003). In addition to Wartegg's (1939, 1954) original work, several test manuals, introductions, and translations have been published from the 1940s to the present day in Germany (Avé-Lallemant, 1978; Lossen & Schott, 1952; Petzold, 1991; Renner, 1953; Vetter, 1952), Italy (Crisi, 1998; Wartegg, 1959, 1972), the United States (Kinget, 1952), Sweden (Wass & Mattlar, 2000), Finland (Gardziella, 1985; Kukkonen, 1962a, 1962b; Takala, 1957), Spain (Wartegg, 1960), France (Fetler-Sapin, 1969; Wartegg, D'Alfonso, & Biedma, 1965), Argentina (Avé-Lallemant, 2001; Biedma & D'Alfonso, 1960), Uruguay (de Gómez & Gómez Pinilla, 1974), Israel (Avé-Lallemant, Vetter, Ben-Asa, & Ari'an-Gafni, 2002), and Indonesia (Kinget, 2000). This indicates that the test is or has been widely known around the world, even though the scope of its use, to our knowledge, has not been verified by any surveys other than those cited here.

Committees evaluating and monitoring psychological test use have recommended against the use of the WZT in Switzerland (Diagnostikkommission des Schweizerischen Verbandes für Berufsberatung, 2004) and in Brazil (Conselho Federal de Psicologia, 2003). However, the Finnish Committee on Psychological Assessment (Testilautakunta, 2008) has seen no reason to ban the use of any method, the WZT included, even though they have admitted that the scientific status of the methods used may vary. A number of academic psychologists in Finland and in Brazil have criticized

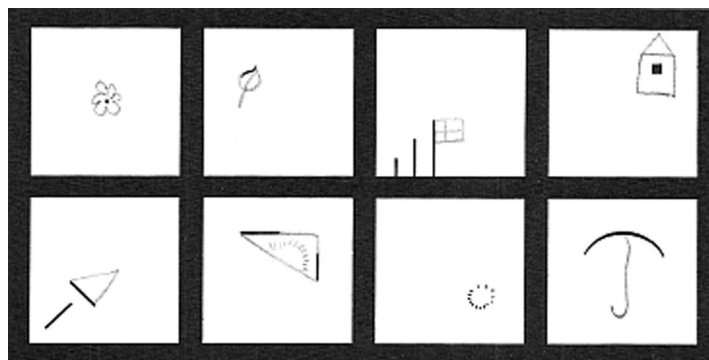


Figure 2. Example of a simple drawing solution (drawn by the first author based on client material). Copyright 1997 Hogrefe Verlag GmbH & Co. KG, Göttingen, Germany. Reprinted with permission.

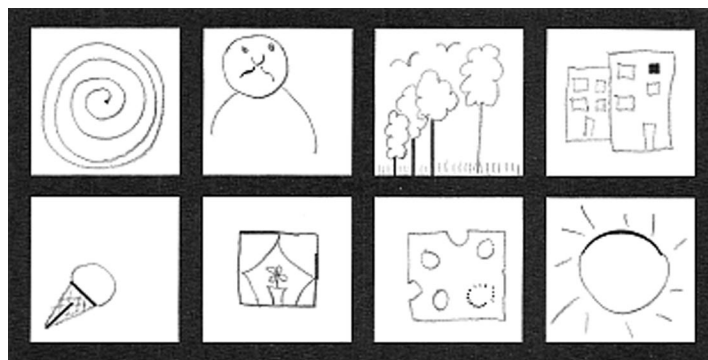


Figure 3. Example of a more complex drawing solution (drawn by the first author). Copyright 1997 Hogrefe Verlag GmbH & Co. KG, Göttingen, Germany. Reprinted with permission.

the use of the WZT in psychological practice, pointing out the lack of proven validity and the extremely small amount of published empirical research (de Souza, Primi, & Miguel, 2007; Nevanlinna, 2004; Noronha, 2002; Nummenmaa & Hyönä, 2005; Ojanen, 1999; Padilha, Noronha, & Fagan, 2007; Pereira et al., 2003; Tamminen & Lindeman, 2000; on the Finnish debate, see Soilevuo Grønnerød & Grønnerød, 2010). Practicing psychologists, on the other hand, have argued for the practical value of the method they feel they experience in their everyday work (Heiska, 2005; Kosonen, 1999).

The small number of empirical studies using the Wartegg method has indeed been noted by several Wartegg authors as well (Kuuskorpi & Keskinen, 2008; Puonti, 2005; Roivainen, 1997, 2006, 2009; Tamminen & Lindeman, 2000). Tamminen and Lindeman (2000, p. 326) assumed that international journals are not interested in publishing articles on a method that they (unfortunately quite incorrectly) claim is used only in Finland. Contrary to these authors, Mattlar (2008) was aware of the scope of Finnish Wartegg research but has pointed out that literature searches on the method have given modest results. When the first author of this article asked a librarian to perform the widest and most complete Wartegg search that was possible in the early 1990s, the result was only a dozen hits. Fifteen years later, Nummenmaa and Hyönä (2005) reported that they had found only 10 hits in PsycINFO. A couple of years later, Roivainen (2009) was able to find 88 hits in PsycINFO, ranging from the 1930s to the 2000s, with a peak in the 1950s.

The Wartegg handbooks cited earlier typically refer to a small number of earlier interpretation manuals, while references to empirical studies are few. Alessandro Crisi, the leader of a Wartegg institute in Italy, has published an online bibliography with 50 entries, including few empirical works (Crisi, 2010). Carl-Erik Mattlar has been working on a literature review on Finnish Wartegg studies and has kindly given us a copy of this as-yet unpublished manuscript with 80 references (Mattlar, 2008). Thus, the problem may not be the complete lack of research on the WZT, but the difficulties in finding the studies.

We argue, first, that the literature studying the Wartegg method has not been easily available because traditions of research and systems of interpretation have developed relatively separately, without cross-references. This separation is partly due to language barriers. The method is almost unknown in the English-speaking

countries, while the WZT finds interested audiences among scholars and practitioners in other parts of the world. Second, Wartegg literature references have escaped the research bibliographies and databases until recently. Earlier literature reviews have been dependent on the authors' knowledge of and physical access to relevant material. The development of online databases and search engines in recent years has now made the whole scope of Wartegg research more visible and available in a way that was not previously possible. It is therefore time to ask: How much has the method actually been studied? Second, based on the available studies, how reliable and valid is the method? To answer these questions, we gathered all the Wartegg method references we could possibly find and used this new bibliography as a base to choose studies for a global meta-analytic review.

## Method

### Literature Overview

We first carried out an extensive literature search in several databases during the month of September 2007. The search term in all databases was *Wartegg*, excluding *hesse* and *schloss*, which turned out to be responsible for most irrelevant hits. We found the following numbers in each search: EMBASE (biomedical and pharmacology database) = 7, ERIC (pedagogy database) = 1, ISI Web of Knowledge (citation index) = 14, ProQuest (dissertation database) = 5, PsycINFO (psychology database) = 91, PubMed (medicine database) = 7, NII-ELS (general database, Japan) = 15, Google Scholar Beta (full-text articles) = 304, and WorldCat Beta (library catalogues worldwide) = 117. The following databases did not return any results: CSA (sociology database), HAPI (health and psychosocial database), NORART (Nordic journals), MedLinePlus (medicine database, 18 hits in a later search). We also conducted searches using the national academic library databases Linda in Finland (42 hits), Bibsys in Norway (4 hits), and Bibliotek.se in Sweden (12 hits). In addition, a regular Google search returned more than 78,000 hits (also excluding *restaurant*), and we checked the first 600 of these. A large portion of the hits in a particular database also appeared in other databases, but they were nonetheless valuable as sources of correction and completion of already existing references. We initially accepted all types of references: journal articles, books, test manuals, academic theses

of all levels, conference papers, unpublished papers, book reviews, and even references we were not able to verify or references that were incomplete. Different versions of the same texts (e.g., translations of books) could be included as separate entries.

Every reasonable effort was made to retrieve as many of the references as possible. We already had a small personal collection of publications gathered over several years to begin with. All available electronic full-text documents were downloaded, and other publications were ordered through the university library. Dissertations from the United States were not ordered, because they would have required extra funding that was not available at the time. Master's theses were not ordered either, because we decided to exclude them from the meta-analysis. Additionally, almost every master's thesis reference was from Finland, which would have made the meta-analysis biased in this respect. Reference lists of the retrieved full-text publications were scanned for additional references not retrieved in the search. Additional searches were made to complete and correct the database up until October 2009.

When we closed the database, it consisted of 507 references, and 238 of these were retrieved in full text. Seven publications were, at this stage, still on request despite several attempts to order them, and they were therefore regarded as unavailable. The first author then classified the 238 full-text publications as a study or not a study. A study was defined as a published text presenting new empirical results in some form (unpublished works, conference papers, and master's theses were thus excluded). All studies were then classified as having one of the following three results: no specific Wartegg method results (e.g., studies mentioning the WZT as one of the methods used but not presenting any specific Wartegg results), no codable results (e.g., qualitative analysis, case studies, only descriptive data), or codable results. Studies written in Japanese could not be translated and had to be regarded as not codable. The number of studies with codable results was 37, which formed the data set for the meta-analysis.

The 37 studies were written in English, Dutch, German, Italian, Portuguese, Finnish, French, and Swedish. Full-text articles in Italian, Portuguese, Dutch, French, or German were fully or partially translated using online translation services (Google Translator at <http://www.translate.google.com> and Bing Translator at <http://www.microsofttranslator.com>) whenever our language skills did not suffice to comprehend the text. Although the automatically translated texts were often grammatically imperfect, the meaning of the text was mostly clear enough for our needs.

## Coding

We coded the 37 studies according to the coding manual presented in the Appendix. The first section included publication year, language, country of origin (defined by first author's affiliation in journal articles or the country of publication in the case of books), publication type, scoring system used, and whether there were reported results. The second section of the coding manual included study design, other assessment methods used in the study, whether scorers were blind to relevant aspects of test origin, and whether reliability checks were performed. For scorer blinding and reliability checks, respectively, we combined Not Reported with either No Blinding or No Reliability Check, because we could not know which was the case when it was not reported. Finally, the third

section included number of subjects in the sample, subject population and age, type of statistic reported, and whether the result was a reliability or a validity coefficient. For validity coefficient, we also coded whether the statistical test was focused; whether the results were in line with specific, clearly formulated hypotheses, contrary to them or merely explorative; and the criteria type applied in the study. In cases where subject age was only reported as adults, we inserted the average age of 38 years, based on the 14 adult samples that reported age ranges or means. This allowed us to perform multiple regression analyses with subject age as one of the variables without excluding the study because of missing data.

To ensure the coding manual provided us with clear and consistent definitions, we calculated intercoder reliability in several stages. First, we double-coded the result coding variable (No Reported Results, No Codable Results, or Codable Results) in 138 studies initially selected as a study by the first author, and the intraclass correlations (ICC) for intercoder reliability was  $ICC(2,1) = .90$ , an excellent level. Next, we came to a consensus on any disagreement on the result variable.

In the next stage, we first jointly coded 18 studies to fine-tune the coding manual definitions. We then selected 20 of the 37 meta-analysis studies to independently double-code according to the whole coding manual. Three studies were excluded for the result statistics calculations because the results were coded by us in a manner that did not allow direct agreement estimation (e.g., one coder had coded one coefficient and the other divided this coefficient into three subcomponents). In all cases, the average of one coder's results equaled the other coder's single or average result, and the coding differences would therefore not affect the overall effect size for the study. All results with this type of disagreement were subsequently coded by consensus.

We calculated one-way random ICCs for scale data, Cramer's V for multicategorical nominal data, and kappa for dichotomous data. In all, we calculated intercoder reliability for 20 variables based on 212 results from 17 studies (excluding the three studies with result coding disagreements) and study data from all 20 studies. The overall average intercoder reliability for our coding was  $r = .84$ . One variable, scorer blinding, resulted in  $ICC(2,1) = -.12$  based on six out of 20 disagreements on whether to code Irrelevant or Not Reported. By excluding this variable, the average reliability was  $r = .95$ . In a final round, all coding disagreements were resolved by consensus coding.

Certain studies were excluded from the meta-analysis for the following reasons. Sisley's (1972, 1973) works were excluded since the studies concern the semantic meanings of the printed stimuli on the Wartegg test blank, thus not presenting empirical results using the method. In six studies, the positive or negative findings could not be readily interpreted as supportive or dismissive of the WZT validity as a personality assessment method. Three of them examined the WZT method's sensitivity to cultural differences (Cuppens, 1969; Forssén, 1979; Gardziella, 1979), Ryhänen et al. (1978) studied possible negative psychological effects of various anesthetic substances, Wassing (1974) investigated an alternate Wartegg form, and Faisal-Cury (2005) compared personality variables in pregnant and not pregnant women.

In three studies, we were not able to determine the correct coding of the validity results (Andreani Dentici, 1994; Ihalaainen, Gardziella, & Hirvenoja, 1973; Mellberg, 1972). In some studies, only significant results were reported, and it was not possible to



determine the number of insignificant results (Brönnimann, 1979; Caricchia, d'Angerio, & Lonoce, 2000; Puonti, 2005). A group of studies in Japanese were, as mentioned, excluded since we were not able to translate them (Daitoku & Nishimura, 2006; Katsura et al., 1974; Sugiura, Hara, Suzuki, Takeuchi, & Kakudate, 2005; Sugiura & Takanashi, 2001; Sugiura & Yagi, 2002; Takayanagi, 2008). For one study, we were only able to retrieve one of its two parts, and the study was therefore not included (Regel, Parnitzke, & Fischel, 1965).

Two studies were challenging because of the large number of their codable results. De Souza, Primi, and Miguel (2007) correlated 141 WZT variables with 16 scales from a personality inventory (16PF), five scales from an intelligence measure, and six scales from an occupational performance assessment questionnaire, leading to a total of 3,807 correlations. Based on the small number of reported significant results in relation to the vast amount of correlations, we opted to code three nonsignificant results, with the WZT correlated against each of the three methods. Takala (1964) was especially challenging in presenting the detailed results of several studies by his colleagues and students, many of them unpublished, in 15 tables of codable results. The number of individual results far outnumbered the number of subjects, and we therefore opted to randomly select four tables for coding, representing each of the topics the author examined: development, intelligence, vocational interest, and correlations with free response methods.

Chi-square coefficients were calculated from figures reported in Teiramaa (1978a, 1978b, 1979a, 1979c) and Crisi (1998) using an online calculator (Preacher, 2010) to get more accurate results. Finally, some results were reported in several publications; therefore Kuha (1981) was coded as Kuha (1973), and Teiramaa (1978b, 1979a, 1979b, 1979c, 1981) were coded as Teiramaa (1978a).

## Meta-Analysis

We coded and calculated the basic meta-analytic effect sizes by entering data and formulas into a spreadsheet. The study effect sizes were then fed into a meta-analysis calculation spreadsheet made by Diener (2009). Although a free spreadsheet, it provided explicit formulas and guidelines for data entry and analysis and has been used in a meta-analysis primer by Diener, Hilsenroth, and Weinberger (2009). We report the random effects model correlational coefficients based on Hedges and Olkin's (1985) procedures presented in this spreadsheet. A random effects model is appropriate rather than a fixed effects model since there is no reason to assume that the studies relate to a common population effect, and we want to make inferences beyond the type of studies included in our analysis (Hedges & Vevea, 1998). We computed study effect sizes as averages of the Fisher-transformed correlation coefficients derived from the individual results, using the spreadsheet formulas for converting different statistics into correlations whenever available. If not available, we entered formulas for the Rosenthal (1991) procedures. All averages were weighted by sample size. Effect sizes were interpreted according to Hemphill's (2003) guidelines.

We ran moderator analyses using correlations, one-way analysis of variance and hierarchical stepwise linear regression procedures in SPSS 16.0. Although strictly a fixed model analysis, we argue that a random model analysis would unduly increase the complex-

ity of the analysis and that the number of studies is too low to estimate error variance for the independent variables. We used the transformed effect sizes as input and the average sample size as weights. Inclusion level for the linear regression was set to  $p = .10$ , and exclusion level to  $p = .15$ . Age was transformed by the log linear function to reduce skewness since adolescents were most frequent and, for substantial reasons, since the significance of, say, a 10-year age difference is much larger in younger age than in older age. The number of coded results was also transformed due to a highly skewed distribution. Both transformations substantially reduced skewness and kurtosis. A dummy variable was created for subject population, dividing them into patients and nonpatients. The criteria variable was recoded into Self-Report, Free Response, Observer, Diagnosis, and Other. We entered eight variables related to samples and method, five of which we entered into the first block of the linear regression analysis: Scorer Blinding, Reliability Check, Subject Age (transformed), Study Design, and Subject Population. The final three we entered into the second block, as we regarded them as being of secondary importance: Publication Year, Publication Type, and Number of Coded Results (transformed). If a variable from the first block was rendered nonsignificant at the entry of the second block, the variable was removed and the analysis rerun.

## Results

### Literature Review

We were able to retrieve 507 references of scholarly works on the WZT, divided into 230 journal publications, 113 books, 40 dissertations, and 124 other types of publications. Among the 31 countries of origin, Germany, Italy, Japan, Brazil, and Finland were the most frequent. We found a clear dip in WZT interest in the 1970s and 1980s, followed by a revival in the 1990s and 2000s.

Compared with earlier accounts on the scope of Wartegg use and research globally, our findings are remarkable. The highest reported number of Wartegg references was 88 in Roivainen (2009). The test history in Germany, Finland, Italy, and Brazil has been documented (Roivainen, 2009), but, to our knowledge, the Japanese Wartegg tradition has not been reported in earlier Wartegg literature at all. The references indicate that the works by Avé-Lallemant (1994), Kinget (1952), and Wartegg (1953) have been used. Other less-known Wartegg traditions have existed in the Netherlands (Evers, Zaal, & Evers, 2002) and in France. Recent translations of test manuals into Hebrew (Avé-Lallemant et al., 2002), Croatian (Kostelić-Martić & Jokić-Begić, 2003), and Indonesian (Kinet, 2000) indicate possible growing traditions in these countries as well.

The isolated nature of the Wartegg traditions is a notable observation. We were surprised to see how few cross-references there are between traditions. In addition, Wartegg studies typically cite only a small number of earlier publications, if any at all, and do not build hypotheses on earlier findings. In our view, this is the result of the invisibility of the literature. It has not been possible to find relevant research, because the works have not been listed in reference databases—not until recently.

## The Meta-Analysis

**Data set.** We based the meta-analysis on 812 individual results from the 37 studies shown in Table 1 ( $N = 7,593$ ). One study (Takala, 1964) reported results from two separate samples, expanding the set to 38 samples. The results consisted of 21 reliability coefficients reported in 15 samples and 791 validity coefficients reported in 33 samples. Both reliability and validity results were reported in nine samples. Thirty of the 791 results were coded as being in the opposite of the expected direction, and 260 as being in the expected direction; the remaining 501 results

were coded as exploratory. This left us with 290 results that we could define as directed based on specific hypotheses, which formed the basis of the main analysis. The contents of the variables covered a wide selection of scoring categories (e.g., drawing order, content categories, pen pressure, and abstraction level), scales and indices (e.g., anxiety index, attachment, and control), and variables based on more informal evaluations (vitality, schizoid self-esteem, extroversion, and sensitivity to stimulus).

**Reliability.** The studies reported three types of reliability results. Interscorer reliability coefficients averaged to  $r_w = .79$  (15 results from 12 samples), a result in the excellent range. The

Table 1  
Sample Coding and Coefficients From 37 Studies

Sample	Country/language	N <sup>a</sup>	Subjects	Moderators <sup>b</sup>	Entries	Coefficient		
						Reliability	Validity	Directed validity
Arajärvi et al. (1974)	Finland/English	151/151	InP	5/0/3/0/0/3/13	1		.29	
Bokslag (1960)	The Netherlands/Dutch	96/96	NPs	5/1/3/0/0/3/16	10		.46	
Brönnimann (1979)	Switzerland/German	190/190/190	NPg	3/12/7/0/0/1/15	1	.45 <sup>c</sup>		
Burbiel & Wagner (1984)	BDR/German	37/37/37	InP	5/12/1/2/2/4/38	23	.68	.37	
Chimenti et al. (1981)	Italy/Italian	390/279/20	NPg	5/12/3/0/2/1/12	6	.89	.17	.35
Crisi (1998)	Italy/Italian	384/372	Som	4/11/2/0/0/0/11	6		.13	
Daini, Bernardini, & Panetta (2007)	Italy/English	91/91	Oth	5/11/4/0/0/2/29	26		.14	.29
Daini, Lai, et al. (2006)	Italy/English	181/181/40	Oth	5/11/4/0/2/5/24	36	.83	.10	
de Caro & Venturino (1991)	Italy/Italian	1,067/1,067	NPg	5/14/2/0/0/3/27	64		.08	.09
de Souza et al. (2007)	Brazil/Portuguese	121/121	NPg	5/13/7/0/0/3/41	3		.06	
Flakowski (1957)	BDR/German	38/38	NPg	5/12/7/0/0/1/11	1		.64	.56
Gardziella (1969)	Finland/Finnish	26/26	NPs	4/8/9/2/0/4/17	4		.34	
Hyypä et al. (1991)	Finland/English	651/651/50	NPg	5/8/5/0/2/4/50	4	.75	.06	.17
Juurmaa & Leskinen (1966)	Finland/Finnish	260/130/50	Som	5/7/2/0/1/0/14	73	.60	.16	
Keith et al. (1966)	USA/English	98/32	NPg	5/4/3/0/0/1/11	18		.10	.11
Konttinen & Olkinuora (1968)	Finland/English	68/68/68	NPg	5/6/7/0/2/2/14	1	.83		
		68/68/68	NPg	5/6/7/0/2/2/14	2	.83 <sup>c</sup>		
Kuha (1973)	Finland/English	150/150	Som	3/4/2/0/0/4/33	7		.09	
Kuha et al. (1975)	Finland/English	100/100	Som	5/4/3/0/0/3/33	7		.14	
Laukkanen (1993)	Finland/Finnish	120/120	OutP	3/8/2/2/0/3/17	25		.19	.37
Markwardt (1961)	DDR/German	52/52	Som	5/0/2/0/0/4/10	1		.09	
Mattlar et al. (1991)	Finland/English	50/50/50	NPg	5/8/8/2/2/0/57	1	.77		
Mellberg (1972)	Finland/English	284/284/284	NPs	5/7/9/0/2/1/15	1	.68		
Pesonen (1970)	Finland/English	127/127	NPs	5/7/9/2/0/1/11	1		.45	.42
Puonti (2005)	Finland/Finnish	29/29/29	NPs	4/10/10/0/2/4/38	2	.91		
		169/169/169	NPs	4/10/10/0/2/4/38	2	.72 <sup>d</sup>		
Roivainen & Ruuska (2005)	Finland/English	83/83/83	OutP	5/12/7/0/2/2/45	4	.94	.27	.33
Scarpellini (1964)	Italy/French	120/120	Oth	5/14/7/0/0/1/22	36		.23	
Silveri et al. (2004)	Italy/English	40/40	Som	5/1/3/2/0/4/76	1		.32	
Soilevuo Grønnerød & Grønnerød (2010)	Norway/English	351/351/50	NPg	5/9/7/0/2/0/20	3	.83	.04	
Takala (1953)	Finland/English	60/60	NPg	5/6/7/0/0/3/22	1		.09	
Takala (1964)								
Children	Finland/English	148/148	NPg	5/6/7/0/0/3/7	60		.10	.10
Adolescents	Finland/English	583/291	NPg	5/6/7/0/0/3/18	168		.12	.11
Takala & Rantanen (1964)	Finland/English	200/200/200	NPg	5/6/7/0/2/2/14	25	.75 <sup>d</sup>	.34	.34
Tamminen & Lindeman (2000)	Finland/Finnish	107/81	NPg	5/8/7/0/1/1/18	5		.12	
Teiramaa (1977)	Finland/English	199/99	Som	3/4/2/2/0/3/34	4		.61	.87
Teiramaa (1978a) <sup>e</sup>	Finland/English	199/145	Som	5/4/10/2/0/3/34	29		.14	.14
Togliatti et al. (2003)	Italy/Italian	389/389	NPg	5/0/2/0/0/2/17	1		.00	
Venturino et al. (1994)	Italy/Italian	843/843	NPg	5/1/2/0/0/2/33	73		.05	
Wass & Mattlar (2000)	Sweden/Swedish	131/87/10	NPg	4/9/4/0/1/1/30	77	.77	.11	

*Note.* All reliability coefficients are interscorer reliability except where noted. InP = Inpatients; NPs = Non-Patients, Selection; NPg = Non-Patients, General; BDR = former West Germany; Som = Somatic Patients; Oth = Other; OutP = Outpatients; DDR = former East Germany.

<sup>a</sup>Total N/average N/reliability coefficient N (if any). <sup>b</sup>Publication Type/Scoring System/Design/Scorer Blinding/Reliability Check/Count of Other Method Types/Average Sample Age. <sup>c</sup>Test-retest reliability. <sup>d</sup>Internal consistency reliability. <sup>e</sup>Teiramaa (1978b, 1979a, 1979b, 1979c, 1981) from the reference list were all coded as Teiramaa (1978a).

calculations were mostly based on single scoring categories covering a wide range of phenomena but also in a few instances on scales based on scoring categories where higher levels should be expected.

Internal consistency coefficients averaged to  $r_w = .74$  (three results from two samples). One study applied a "split-half reliability" procedure without specifying how the split was done, and the other study calculated internal consistency for scales based on WZT scoring categories. The levels were satisfactory, although it is doubtful whether split-half reliability is relevant for the WZT given the unique character of each square.

Test-retest reliability, on the other hand, was disappointingly low, with a weighted average of  $r_w = .53$  (three results from two samples within 1 week and 1- to 3-week retest periods). This is especially true given the short retest periods. The results were calculated on the basis of both content scores and scores related to drawing style. Given the lack of empirical support for specific variables, it is difficult to conclude whether the variations in level can be attributed to state- or trait-related characteristics. The weighted average reliability coefficient for all three reliability types was  $r_w = .74$ .

**Validity.** The results from the random effects model was based on 290 results coded in 14 studies ( $N = 3,693$ ) as being based on a clear and specific hypothesis. The model yielded  $r_w = .33$ , a large effect size (95% confidence interval [CI: .18, .47]; heterogeneity  $Q = 304.18$ ,  $p < .0000$ ). Clearly, defining a good rationale and specifying hypotheses is related to powerful results.

The weighted average validity coefficient for all results was  $r_w = .19$ , a lower middle magnitude effect size (95% CI [.14, .26]). The significant heterogeneity ( $Q = 225.87$ ,  $p < .0000$ ) cautioned us not to interpret the levels directly but rather to investigate further the influences that various factors had on the levels.

Second, we examined different criteria used in the studies. Fourteen samples used more than one criterion, and we calculated one effect size for each. This resulted in a list of  $k = 51$  effect sizes to be compared. The overall differences between criteria, weighted by sample size, was significantly different,  $F(4, 9942) = 417.149$ ,  $p < .000$  (averages shown in Table 2). Contrast analyses showed that the difference between self-report and free response criteria was significantly different,  $t(9942) = 21.678$ ,  $p < .000$ , and the same was true for free response and observer compared with self-report, diagnosis, and other,  $t(9942) = 22.916$ ,  $p < .000$ .

Third, we grouped samples according to scoring system used. The systems differed significantly,  $F(7, 6162) = 325.431$ ,  $p < .000$  (averages shown in Table 2). Interpretation is somewhat more difficult, however, since only a few studies represented each system.

Fourth, the regression model identified important variables related to variances in effect size levels, as shown in Table 3. Studies that employed and reported scorer blinding reported, on average, larger effect sizes than did those that ignored scorer blinding. The other moderator was publication year, showing a decline in effect sizes over the 79-year period covered. The second step model produced rather strange predictions, however (e.g., effect size for publication year 1939 was .66), and we opted to use the first model. By entering No Scorer Blinding into the equation, the model predicted an effect size of  $r = .12$ , whereas Full Scorer Blinding predicted a level of  $r = .35$ .

Table 2  
Sample Effect Sizes by Criteria ( $k = 51$ ) and Scoring System ( $k = 30$ )

Subdivision	$k$	$r_w$
Criteria		
Self-report	13	.15
Free response	5	.28
Observer	7	.20
Diagnosis	12	.10
Other	14	.14
Scoring system		
Wartegg and others <sup>a</sup>	5	.10
Gardziella and others <sup>b</sup>	6	.08
Kinget (1952)	5	.23
Crisi (1998)	3	.13
South American systems <sup>c</sup>	1	.06
Takala (1957)	4	.18
Kukkonen (1962a, 1962b)	2	.31
Several scoring systems	4	.26

Note.  $k$  = number of samples;  $r_w$  = average effect size weighted by sample  $N$ .

<sup>a</sup> Avé-Lallemant (1978), Lossen and Schott (1952), Renner (1953), Vetter (1952), and Wartegg (1939, 1953). <sup>b</sup> Gardziella (1985), Puonti (2005), and Wass and Mattlar (2000). <sup>c</sup> Biedma and D'Alfonso (1960) and Freitas (1993).

## Discussion

When planning this meta-analytic study, we were aware of the debate on the Wartegg method and of the uncompromising positions of the opposite camps. We did not have any preference for the results of the analysis—we would have been equally satisfied with any result, either supporting the validity of the method or not. In fact, a clear result showing poor validity results would have been easier to communicate to the scientific community. Also, in the course of the coding process, our skepticism grew because many studies did not report scorer blinding or interscorer reliability, and clear hypotheses were presented all too seldom.

We were thus surprised to find an effect size of  $r_w = .33$ , a relatively strong result, when looking at studies testing specific hypotheses. Clearly, too many WZT studies have not been conducted adequately, and we see a larger variation in WZT study quality than in studies of other widespread methods. The Hiller, Rosenthal, Bornstein, Berry, and Brunell-Neuleib (1999) meta-analysis of the Rorschach and the MMPI applied inclusion criteria that resemble our focus on specific hypotheses, and their levels of .29 for the Rorschach and .30 for the MMPI are quite comparable. Meta-analytic studies of the Rorschach method, the MMPI (Butcher, Dahlstrøm, Graham, Tellegen, & Kraemmer, 1989), and the Thematic Apperception Test (Murray, 1943) have generally reported levels around .30 (Garb, Florio, & Grove, 1998; Hiller et al., 1999; Meyer & Archer, 2001; Rosenthal, Hiller, Bornstein, & Berry, 2001; but see Meyer, 2004, for a review of various levels). We conclude, therefore, that research on the WZT may reach levels comparable to other assessment methods, given sufficient focus on study quality.

The most important recent critiques of the validity of the Wartegg method have been presented by Tamminen and Lindeman (2000) and de Souza et al. (2007). Tamminen and Lindeman (2000) studied WZT validity against four subscales of the Person-

Table 3  
Weighted Regression Model of Moderator Influences

Moderator	R	Constant	r <sup>a</sup>	B	SE	β
Step 1	.513	0.116			0.022	
Scorer blinding			.45**	0.122	0.037	.513***
Step 2	.679	8.335			2.480	
Scorer blinding			.45**	0.107	0.032	.449***
Publication year			.37*	0.004	0.001	.449***

<sup>a</sup> Two-tailed univariate correlation with untransformed effect size.

\*  $p < .05$ . \*\*  $p < .005$ . \*\*\*  $p < .001$ .

ality Research Form (Jackson, 1974), State-Trait Anxiety Test (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1977), and Adult Attachment Styles (Hazan & Shaver, 1987) in a student sample, finding no support for interpretations according to Gardziella's (1985) system. The de Souza et al. (2007) study strongly criticized the WZT method by comparing it with the 16 PF (Cattell & Cattell, 1995). Even if the lack of correlations between the WZT and the self-report measures may look convincing, one must keep in mind the general lack of relationship between free response and self-report methods, also termed the *heteromethod convergence problem* (Bornstein, 2009). Also in our data, the WZT method was more strongly related to other free response methods and observations than to self-report methods.

Our aim was to link the meta-analysis to the historical developments of the Wartegg method in order to understand the specific problems of the method. Small-scale research traditions have been active and vanished in various parts of the world, and links between them have been few and occasional, thus creating a sporadic research tradition. Studies using the Wartegg method typically have not referred to relevant earlier results, either because such studies did not exist or the information on them was not available. In addition, different traditions have used different scoring systems and related to different personality characteristics. Research on the Wartegg method has thus failed to produce cumulative knowledge. We hope that this article will help to create networks between researchers working on the WZT. The Wartegg bibliography (Soilevuo Grønnerød & Grønnerød, 2011) we gathered will be available from our website (<http://wartegg.info>).

The history of the Wartegg method reflects geographical changes in modern psychology. The case of Peru, described by de Rueda (2002), possibly applies for similar countries as well. The method was introduced in Peru as a part of the strong German influence on the development of Peruvian psychology and psychiatry at the beginning of the 20th century. After World War II and during the 1950s, the German roots of Peruvian psychology were set aside. In global political history, the WZT found itself in the wrong place at the wrong time, and an association with the Nazi regime was inevitable. Ehrig Wartegg worked in Leipzig in the former East Germany with limited possibilities for developing the method and little help for increasing its use (Klemperer, 2000; Lockot, 2000; Roivainen, 2009). In Finland and Sweden, the development and dissemination of knowledge on the method has been dependent on personal contacts, and we think it is reasonable to assume that this is also the case in other countries. The WZT also suffered from a language barrier as the English-language publications took over and dominated the library databases until recent years.

In conclusion, based on our meta-analysis, we argue that there is no reason to dismiss the Wartegg method altogether as a method for personality evaluation. However, it is necessary to build a solid, cumulative research tradition to produce knowledge and create a basis for the use of the Wartegg method in psychological practice. As a method that is easy to administer and not entirely dependent on language (although cultural differences have been reported; see Cuppens, 1969; Forssén, 1979; Gardziella, 1979), the Wartegg method may become a useful addition to a practicing psychologist's toolkit. For the time being, we call for caution when using the method as a basis for critical decisions in psychological practice. We strongly encourage, however, more research built on previous studies that will cultivate the strongest part of the method.

## References

References marked with an asterisk indicate studies that are included in the meta-analysis.

- Andreani Dentici, O. (1994). Pensiero logico e immaginazione negli adolescenti [Logical thinking and imagination in adolescents]. *Archivio di Psicologia, Neurologia e Psichiatria*, 55, 192–219.
- \*Arajärvi, T., Mäkönen, K., Repo, I., & Torma, S. (1974). On specific reading and writing difficulties of pre-adolescents. *Psychiatria Fennica*, 231–236.
- Avé-Lallemant, U. (1978). *Der Wartegg-Zeichentest in der Jugendberatung: Mit systematischer Grundlegung von August Vetter* [The Wartegg Drawing Completion Test in youth counseling: With systematic foundations by August Vetter]. Munich, Germany: Ernst Reinhardt Verlag.
- Avé-Lallemant, U. (1994). *Der Wartegg-Zeichentest in der Lebensberatung: Mit systematischer Grundlegung von August Vetter* [The Wartegg Drawing Completion Test in counseling: With systematic foundations by August Vetter] (2nd ed.). Munich, Germany: Ernst Reinhardt Verlag.
- Avé-Lallemant, U. (2001). *El test de dibujos Wartegg: Su aplicación en niños, adolescentes y adultos* [The Wartegg Drawing Completion Test: Its application on children, adolescents and adults]. Buenos Aires, Argentina: Lasra Ediciones.
- Avé-Lallemant, U., Vetter, A., Ben-Asa, M., & Ari'an-Gafni, T. (2002). *Mivhan Varteg: Keli le-ivhun ve-yi'uts* [The Wartegg Test: An instrument for diagnostics and counseling]. Jerusalem, Israel: Keter.
- Beck, S. J. (1930). The Rorschach Test and personality diagnosis: I. The feeble-minded. *American Journal of Psychiatry*, 87, 19–52.
- Biedma, C. J., & D'Alfonso, P. G. (1960). *El lenguaje del dibujo: Test de Wartegg-Biedma-D'Alfonso (versión modificada del test de Wartegg, traducción de Shuji Murata)* [The language of drawing: The Wartegg-Biedma-D'Alfonso Test (modified version of the Wartegg Drawing Completion Test, translation by Shuji Murata)]. Buenos Aires, Argentina: Editorial Kapelusz.
- \*Bokslag, J. G. H. (1960). De predictieve waarde van de tekentest van Wartegg (W. Z. T.) en van een verrichtingstest (Block Design uit Wechsler-Bellevue-Scale) bij de selectie van leerjongens voor drie bed-



- rijven. [The predictive value of the Wartegg drawing test and of a performance test (Wechsler Block Design) in the selection of apprentices for three industrial plants]. *Nederlandsch Tijdschrift voor Psychologie*, 15, 321–333.
- Bornstein, R. F. (2009). Heisenberg, Kandinsky, and the heteromethod convergence problem: Lessons from within and beyond psychology. *Journal of Personality Assessment*, 91, 1–8. doi:10.1080/00223890802483235
- \*Brönnimann, M. (1979). *Beziehungen zwischen dem Wartegg-Zeichentest (WZT) und dem Deutschen High School Personality Questionnaire (HSPQ) von Schumacher/Cattell* [Relationships between the Wartegg Drawing Completion Test and the German High School Personality Questionnaire (HSPQ) by Schumacher/Cattell]. Bern, Switzerland: Peter Lang AG.
- \*Burbiel, I., & Wagner, H. (1984). Einige Ergebnisse dynamisch-psychiatrischer Effizienzforschung. [Some results of dynamic-psychiatric efficiency research]. *Dynamische Psychiatrie*, 17, 468–500.
- Butcher, J. N., Dahlström, W. G., Graham, J. R., Tellegen, A., & Kraemer, B. (1989). *Manual for the Restandardized Minnesota Multiphasic Personality Inventory: MMPI-2: An administration and interpretive guide*. Minneapolis, MN: University of Minnesota Press.
- Caricchia, F., d'Angerio, S., & Lonoce, G. (2000). Il bambino in età prescolare all'esame del test di Wartegg [The preschool child examined by the Wartegg Drawing Completion Test]. *Babele*, No. 14.
- Cattell, R. B., & Cattell, H. E. P. (1995). Personality structure and the new fifth edition of the 16PF. *Educational and Psychological Measurement*, 55, 926–937.
- Ceccarelli, C. (1999). *L'Uso degli strumenti psicodiagnostici* [The use of psychodiagnostic methods]. Retrieved May 31, 2007, from the Società Italiana di Psicologia dei Servizi Ospedalieri e Territoriali website: <http://www.sipsot.it>
- \*Chimenti, R., de Coro, A., & Grasso, M. (1981). Proposta di alcune scale empiriche di valutazione per test grafici: Contributo allo studio dell'identificazione e differenziazione sessuale in preadolescenza e adolescenza. [Proposal of some empirical evaluation scales for graphic tests: A contribution to the study of sexual identification and differentiation in preadolescence and adolescence]. *Bollettino di Psicologia Applicata*, 159, 83–116.
- Conselho Federal de Psicologia. (2003). *Edital CFP N. 2 de 6.11.2003: Processo de avaliação dos testes psicológicos*. [Announcement CFP No. 2, November 6, 2003: Assessment of psychological tests]. Retrieved September 12, 2007, from [http://www.pol.org.br/servicos/pdf/editalcfp\\_testespsi\\_n2.pdf](http://www.pol.org.br/servicos/pdf/editalcfp_testespsi_n2.pdf)
- \*Crisi, A. (1998). *Manuale del Test di Wartegg* [Wartegg Test Manual]. Rome, Italy: Edizioni Scientifiche Ma. Gi. srl.
- Crisi, A. (1999, July). *A new methodology for the clinical use of the Wartegg test*. Paper presented at the 16th Congress of the International Rorschach Society, Amsterdam, the Netherlands.
- Crisi, A. (2008, March). *A new instrument for selection and career guidance within the armed forces: The Wartegg Test*. Paper presented at the Society for Personality Assessment Annual Meeting, New Orleans, Louisiana.
- Crisi, A. (2010). *Bibliografia relativa al Test di Wartegg* [Wartegg Test bibliography]. Retrieved May 19, 2010, from <http://wartegg.com/bibliografia.php>
- Cuppens, E. Ch. J. (1969). De Wartegg-Teken-Test: Evaluatie van een omstreden techniek. [The Wartegg Drawing Completion Test: Evaluation of a controversial technique]. *Gawain*, 17, 1–70.
- \*Daini, S., Bernardini, L., & Panetta, C. (2007). A personality perspective on female infertility: An analysis through Wartegg test. *Journal of Projective Psychology & Mental Health*, 14, 135–144.
- \*Daini, S., Lai, C., Festa, G. M., Maiorino, F., Pertosa, M., & De Risio, S. (2006). Impulsivity in eating disorders: Analysed through Wartegg test. *Journal of Projective Psychology & Mental Health*, 13, 107–117.
- Daitoku, R., & Nishimura, Y. (2006). A research of the development of children with slight developmental disorders through drawing test, using Star-Wave and Wartegg-Zeichen Test. *Journal of Nishikyushu University & Saga Junior College*, 36, 59–69.
- \*de Caro, R., & Venturino, G. (1991). Test del disegno di Wartegg: Primi dati di un'indagine condotta su un campione ristretto di una popolazione studentesca ad estrazione artistica [The Wartegg Drawing Test: First results of a survey conducted on a small sample of art students]. *Neurologia Psichiatria Scienze Umane*, 11, 489–507.
- de Godoy, S. L., & Noronha, A. P. P. (2005). Instrumentos psicológicos utilizados em seleção profissional [Psychological instruments of recruiting process]. *Revista do Departamento de Psicologia UFF*, 17, 139–159.
- de Gómez, M. L. M., & Gómez Pinilla, J. C. (1974). *Test Wartegg-color* [The Wartegg Color Test]. Montevideo, Uruguay: Editorial Losada Uruguay.
- Deinlein, W., & Boss, M. (2003). *Befragung zum Stand der Testanwendung in der allgemeinen Berufs- Studien- und Laufbahnbereitung in der deutschen Schweiz* [The use of tests in vocational counseling in German-speaking Switzerland]. Dissertation Nachlizentiatstudium in Occupation, Studies and Career Consultation NABB-6.
- de Oliveira, K. L., Noronha, A. P. P., Dantas, M. A., & Santarem, E. M. (2005). O psicólogo comportamental e a utilização de técnicas e instrumentos psicológicos [The use of psychological techniques and instruments for behavioral psychologists]. *Psicologia em Estudo*, 10, 127–135. doi:10.1590/S1413-73722005000100015
- de Rueda, M. d. C. B. (2002). Saat und ernte—Deutsche psychiatrie und psychologie in Perú [Seed and harvest—German psychiatry and psychology in Peru]. *Fortschritte der Neurologie—Psychiatrie*, 70, 259–267.
- \*de Souza, C. V. R., Primi, R., & Miguel, F. K. (2007). Validade do Teste Wartegg: Correlação com 16PF, BPR-5 e Desempenho Profissional [The validity of the Wartegg Drawing Completion Test: Correlations with 16PF, BPR-5 and a professional performance measure]. *Avaliação Psicológica*, 6, 39–49.
- Diagnostikkommission des Schweizerischen Verbandes für Berufsberatung. (2004). *Label des Wartegg Zeichentests* [Ratification of the Wartegg Drawing Test]. Retrieved September 13, 2007, from <http://www.testraum.ch/Serie%207/wzt.htm>
- Diener, M. (2009). *Meta-analysis of correlation coefficients* [Calculation spreadsheet]. Retrieved September 28, 2009, from <http://www.informaworld.com/mpp/uploads/metaanalysisprogramv.3.4.xls>
- Diener, M. J., Hilsenroth, M. J., & Weinberger, J. (2009). A primer on meta-analysis of correlation coefficients: The relationship between patient-reported therapeutic alliance and adult attachment style as an illustration. *Psychotherapy Research*, 19, 519–526. doi:10.1080/10503300802491410
- Evers, A., Zaal, J. N., & Evers, A. K. (2002). Ontwikkelingen in het testgebruik van Nederlandse psychologen [Developments in test use by Dutch psychologists]. *Psycholoog Amsterdam*, 37, 54–61.
- Faisal-Cury, A. (2005). Características psicológicas da primigestação [Psychological characteristics of the first pregnancy]. *Psicologia em Estudo*, 10, 383–391. doi:10.1590/S1413-73722005000300006
- Fetler-Sapin, M. E. (1969). Une épreuve d'expression graphique: Le test de Wartegg [A test of graphic expression: The Wartegg Test]. *Revue de Psychologie et des Sciences de l'Éducation*, 4, 437–447.
- \*Flakowski, H. (1957). Entwicklungsbedingte Stilformen von Kinderaufsatz und Kinderzeichnung [Developmentally conditioned styles of children's compositions and drawings]. *Psychologische Beiträge*, 3, 446–467.
- Forssén, A. (Ed.). (1979). *Roots of traditional personality development among the Zaramo in coastal Tanzania*. Mikkeli, Finland: Central Union for Child Welfare in Finland.
- Freitas, A. M. L. (1993). *Guia de Aplicação e Avaliação do Teste Wartegg*

- [Test manual for Wartegg assessment]. Sao Paulo, Brazil: Casa do Psicólogo.
- Garb, H. N., Florio, C. M., & Grove, W. M. (1998). The validity of the Rorschach and the Minnesota Multiphasic Personality Inventory: Results from meta-analyses. *Psychological Science*, 9, 402–404. doi:10.1111/1467-9280.00075
- \*Gardziella, M. (1969). Wartegg-testin validiteetti [The validity of the Wartegg Test]. In M. Jääskeläinen & A. Valpola (Eds.), *Ammattikoulumenestyksen ennustaminen* [Prediction of vocational school performance] (pp. 59–66). Helsinki, Finland: Työterveyslaitoksen tutkimuksia 48.
- Gardziella, M. (1979). Gedanken zu Warteggtest-Zeichnungen von Kindern des Zaramo-stammes [Thoughts about Wartegg Drawing Completion Test drawings from children in the Zaramo tribe]. In A. Forssén (Ed.), *Roots of traditional personality development among the Zaramo in coastal Tanzania* (pp. 69–87). Mikkeli, Finland: Central Union for Child Welfare in Finland.
- Gardziella, M. (1985). *Wartegg-piirustustesti: Käsikirja* [The Wartegg Drawing Test: A handbook]. Jyväskylä, Finland: Psykologien Kustannus Oy.
- Hazan, C., & Shaver, P. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology*, 52, 511–524. doi:10.1037/0022-3514.52.3.511
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3, 486–504. doi:10.1037/1082-989X.3.4.486
- Heiska, J. (2005). Projektiiviset testit—kyllä vai ei [Projective methods—yes or no]. *Psykologi*, 23.
- Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, 58, 78–79. doi:10.1037/0003-066X.58.1.78
- Hertz, M. (1935). Rorschach norms for an adolescent age group. *Child Development*, 6, 69–76. doi:10.2307/1125557
- Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T. R., & Brunell-Neuleib, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment*, 11, 278–296. doi:10.1037/1040-3590.11.3.278
- \*Hyypä, M. T., Kronholm, E., & Mattlar, C.-E. (1991). Mental well-being of good sleepers in a random-population sample. *British Journal of Medical Psychology*, 64, 25–34. doi:10.1111/j.2044-8341.1991.tb01639.x
- Ihalainen, O., Gardziella, M., & Hirvenoja, R. (1973). Winter bathing. *Psychiatria Fennica*, 71–77.
- Jackson, D. N. (1974). *Personality Research Form: Manual*. Port Huron, MI: Research Psychologists Press.
- \*Juurmaa, J., & Leskinen, M. (1966). Kuuloalueen toimintojen puutumisen tai olennaisen heikkenemisen vaikutuksesta persoonallisuuden piirteisiin: Eräiden hypoteesien johtoa ja Wartegg-testin ilmaisuihin perustuvaa empiiristä tarkastelua [The effect of hearing deficiencies on personality: Hypotheses and an empirical analysis of the Wartegg results]. *Acta Psychologica Fennica*, 2, 107–132.
- Katsura, T., Saito, K., Tahara, M., Yamada, T., Morishita, A., Oishi, M., ... Maki, S. (1974). A case of psychogenic vegetosis with marked improvement following psychotherapy. *Journal of the Japanese Psychosomatic Society*, 14, 253–257.
- \*Keith, J. P., Jordan, J. E., & Matheny, K. B. (1966). A cross-cultural study of potential school dropouts in certain sub-Saharan countries. *Journal of Negro Education*, 35, 90–94. doi:10.2307/2293935
- Kinget, G. M. (1952). *The Drawing-Completion Test: A projective technique for the investigation of personality*. New York, NY: Grune & Stratton.
- Kinget, G. M. (2000). *Wartegg: Tes melengkapi gambar* [Wartegg: Complete the test image]. Yogyakarta, Indonesia: Pustaka Pelajar.
- Klemperer, P. (2000). Eine Einstimmung auf Ehrig Warteggs autobiographische Skizze: “Zeichen der Zeit” [A match for Ehrig Wartegg’s autobiographical sketch: “Drawings of time”]. In H. Bernhardt & R. Lockot (Eds.), *Mit Ohne Freud: Zur Geschichte der Psychoanalyse in Ostdeutschland* [Without Freud: On the history of psychoanalysis in East Germany] (pp. 92–94). Giessen, Germany: Psychosozial-Verlag.
- \*Kontinen, R., & Olkinuora, E. (1968). Generality of graphic variables across drawing tasks. *Scandinavian Journal of Psychology*, 9, 161–168. doi:10.1111/j.1467-9450.1968.tb00531.x
- Kosonen, P. (1999). Tieteen papit ja käytännön työvälineet, eli miksi Wartegg toimii? [The high priests of science and the tools of practice: Why does the Wartegg work?]. *Psykologi*, 30–31.
- Kostelić-Martić, A., & Jokić-Begić, N. (2003). *Priručnik: Wartegg test crteža* [Manual: Wartegg Completion Test drawings]. Zagreb, Croatia: Školska knjiga.
- \*Kuha, S. (1973). *A psychosomatic approach to pulmonary tuberculosis* (Acta Universitatis Ouluensis, Series D, Medica No. 3, Psychiatrica No. 2). Oulu, Finland: University of Oulu.
- Kuha, S. (1981). Psychosocial factors in the development of pulmonary tuberculosis. *Psychiatria Fennica* (Suppl.), 79–84.
- \*Kuha, S., Moilanen, P., & Kampman, R. (1975). The effect of social class on psychiatric psychological evaluations in patients with pulmonary tuberculosis. *Acta Psychiatrica Scandinavica*, 51, 249–256.
- Kukkonen, S. (1962a). WZT-käsikirja I: Variaabelien pisteistys- ja tulkin-taohjeet [WZT handbook I: Scoring and interpretation]. Helsinki, Finland: Kulkulaitosten ja yleisten töiden ministeriö, Ammatinvalinnano-hjaustoimisto.
- Kukkonen, S. (1962b). WZT-käsikirja II: Havaintomateriaali [WZT hand-book II: Examples]. Helsinki, Finland: Kulkulaitosten ja yleisten töiden ministeriö, Ammatinvalinnano-hjaustoimisto.
- Kuuskorpi, T., & Keskinen, E. (2008). *Psykologisten testien käyttö Suomessa: Testaamisen määrä ja yleisimmät testit* [The use of psycho-logical tests in Finland: The frequency of testing and the most frequently used tests]. Retrieved April 10, 2008, from <http://www.testilautakunta.fi/Artikkeli.pdf>
- \*Laukkanen, E. (1993). *Nuoruusiän psyykinen kehitys ja sen häiriintymi-nen* [The psychic development in adolescence and related disorders]. Kuopio, Finland: University of Kuopio.
- Lockot, R. (2000). Ehrig Warteggs Selbstverwirklichung in der Andeu-tung: Kurt Höck über seinen langjährigen Mitarbeiter Ehrig Wartegg. [Ehrig Wartegg’s self-realization in the suggestion: Kurt Höck on his long-standing coworker Ehrig Wartegg]. In H. Bernhardt & R. Lockot (Eds.), *Mit Ohne Freud: Zur Geschichte der Psychoanalyse in Ost-deutschland* [Without Freud: On the history of psychoanalysis in East Germany] (pp. 118–127). Giessen, Germany: Psychosozial-Verlag.
- Lossen, H., & Schott, G. (1952). *Gestaltung und Verlaufsdynamik: Ver-such einer prozessualen Analyse des Warteggzeichentestes* [Gestalt and flow dynamics: An attempt to form a processual analysis of the Wartegg Drawing Completion Test]. Biel, Germany: Institut für Psycho-Hygiene.
- \*Markwardt, A. M. (1961). Vorläufige Erfahrungen über die Auswirkung der Gaumennahterweiterung auf das Hilfsschulkind [Preliminary expe-riences with the effects of palate extension on auxiliary school children]. *Fortschritte der Kieferorthopädie*, 22, 359–364. doi:10.1007/BF02165928
- Mattlar, C.-E. (2008). *The Wartegg Zeichen Test (WZT): An overview of the method, and of research supporting its utility*. Unpublished manu-script.
- \*Mattlar, C.-E., Lindholm, T., Haasiosalo, A., Vesala, P., Rissanen, S., Santasalo, H., ... Puukka, P. (1991). Interrater agreement when assess-ing alexithymia using the Drawing Completion Test (Wartegg Zeichen-test). *Psychotherapy and Psychosomatics*, 56, 98–101. doi:10.1159/000288538

- \*Mellberg, K. (1972). The Wartegg Drawing Completion Test as a predictor of adjustment and success in industrial school. *Scandinavian Journal of Psychology*, 13, 34–38. doi:10.1111/j.1467-9450.1972.tb00046.x
- Meyer, G. J. (2004). The reliability and validity of the Rorschach and Thematic Apperception Test (TAT) compared with other psychological and medical procedures: An analysis of systematically gathered evidence. In M. J. Hilsenroth & D. L. Segal (Eds.), *Comprehensive Handbook of Psychological Assessment: Vol. 2. Personality assessment* (pp. 315–342). Hoboken, NJ: Wiley.
- Meyer, G. J., & Archer, R. P. (2001). The hard science of Rorschach research: What do we know and where do we go? *Psychological Assessment*, 13, 486–502. doi:10.1037/1040-3590.13.4.486
- Meyer, G. J., & Kurtz, J. E. (2006). Advancing personality assessment terminology: Time to retire “objective” and “projective” as personality test descriptors. *Journal of Personality Assessment*, 87, 223–225. doi:10.1207/s15327752jpa8703\_01
- Murray, H. A. (1943). *Thematic Apperception Test: Manual*. Cambridge, MA: Harvard University Press.
- Nevanlinna, R. (2004, February 6). P-koe ennustaa oikein [The P-test predicts correctly]. *Ruotuväki*, No. 3.
- Noronha, A. P. P. (2002). Os problemas mais graves e mais frequentes no uso dos testes psicológicos [The worst and the most common problems in the use of the psychological tests]. *Psicologia: Reflexão e Crítica*, 15, 135–142. doi:10.1590/S0102-79722002000100015
- Noronha, A. P. P., Primi, R., & Alchieri, J. C. (2005). Instrumentos de avaliação mais conhecidos/ utilizados por psicólogos e estudantes de psicologia [The most used assessment instruments by psychologists and psychology students]. *Psicologia: Reflexão e Crítica*, 18, 390–401. doi:10.1590/S0102-79722005000300013
- Nummenmaa, L., & Hyönä, J. (2005). Voiko projektiviisiin testeihin luottaa? [Can we trust projective tests?] *Psykologi*, 14–16.
- Ojanen, M. (1999). Projektivistien testien luotettavuus: *Kertovatto musteläiskät tai käsiala jotain ihmisen persoonallisuudesta?* [The validity of the projective tests: Do ink blots or handwriting tell something about personality?]. *Skeptikko*, 1, 4–10.
- Padilha, S., Noronha, A. P. P., & Fagan, C. Z. (2007). Instrumentos de avaliação psicológica: Uso e parecer de psicólogos [Instruments for psychological assessment: Use and evaluation by psychologists]. *Avaliação Psicológica*, 6, 69–76.
- Pereira, F. M., Primi, R., & Cobêro, C. (2003). Validade de testes utilizados em seleção de pessoal segundo recrutadores [Validity of personal selection tests according to professionals]. *Psicologia: Teoria e Prática*, 5, 83–98.
- \*Pesonen, J. (1970). *Practical application of the Wartegg test in the prediction of the school adjustment of elementary school boys* (Research Report No. 2). Jyväskylä, Finland: University of Jyväskylä, Department of Special Education.
- Petzold, H. (1991). *Der Wartegg-Zeichentest (WZT): Einführung und Auswertungsrichtlinien* [The Wartegg Drawing Test (WZT): Application and interpretation]. Berlin, Germany: Self-published.
- Preacher, K. J. (2010). *Calculation for the chi-square test: An interactive calculation tool for chi-square tests of goodness of fit and independence*. Retrieved April 8, 2010, from <http://www.quantpsy.org>
- \*Puonti, M. (2005). *NBS—Need Based System: Käsikirja* [A handbook]. Helsinki, Finland: Improvia Oy.
- Regel, H., Parnitzke, K. H., & Fischel, W. (1965). Die Verwendung graphischer Verfahren bei der Diagnostik frühkindlicher Hirnschädigungen [The application of graphical methods in diagnosing early cerebral lesions]. *Acta Paedopsychiatrica*, 32, 338–343.
- Renner, M. (1953). *Der Wartegg-Zeichentest im Dienste der Erziehungsberatung: Nach der Auswertung von Vetter* [The Wartegg Drawing Test in the service of educational counseling]. Munich, Germany: Ernst Reinhardt Verlag.
- Roivainen, E. (1997). *Onko Wartegg-piirustustesti validi?* [Is the Wartegg valid?]. Unpublished manuscript.
- Roivainen, E. (2006). Ehrig Wartegg ja Wartegg-testin varhaisvaiheet [Ehrig Wartegg and the early history of Wartegg's Drawing Test]. *Psykologia*, 41, 260–268.
- Roivainen, E. (2009). A brief history of the Wartegg Drawing Test. *Gestalt Theory*, 31, 55–71.
- \*Roivainen, E., & Ruuska, P. (2005). The use of projective drawings to assess alexithymia: The validity of the Wartegg Test. *European Journal of Psychological Assessment*, 21, 199–201. doi:10.1027/1015-5759.21.3.199
- Rosenthal, R. (1991). *Meta-analytic procedures in social research* (Rev. ed.). Baltimore, MD: Sage.
- Rosenthal, R., Hiller, J. B., Bornstein, R. F., & Berry, D. T. R. (2001). Meta-analytic methods, the Rorschach, and the MMPI. *Psychological Assessment*, 13, 449–451. doi:10.1037/1040-3590.13.4.449
- Ryhänen, P., Helkala, E. L., Ihala, O., Hollmén, A., Rantakylä, S., Merilä, M., . . . Hörttonen, L. (1978). Effects of anaesthesia on the psychological function of patients. *Annals of Clinical Research*, 10, 318–322.
- \*Scarpellini, C. (1964). Diagnostic différentiel à travers deux tests de personnalité: Le Rorschach et le Wartegg [Differential diagnostic through two personality tests: The Rorschach and the Wartegg]. *Bulletin de Psychologie Scolaire et d'Orientation*, 13, 79–92.
- \*Silveri, M. C., Salvigni, B. L., Jenner, C., & Colamónico, P. (2004). Behavior in degenerative dementias: Mood disorders, psychotic symptoms and predictive value of neuropsychological deficits. *Archives of Gerontology and Geriatrics*, 38, 365–378. doi:10.1016/j.archger.2004.04.047
- Sisley, E. L. (1972). Differential perception of Kinget's drawing-completion stimuli. *Perceptual and Motor Skills*, 35, 491–494.
- Sisley, E. L. (1973). The meaning of drawing-completion stimuli: New guidelines for projective interpretation. *Journal of Personality Assessment*, 37, 64–68. doi:10.1080/00223891.1973.10119830
- Society for Personality Assessment. (2005). The status of the Rorschach in clinical and forensic practice: An official statement by the Board of Trustees of the Society for Personality Assessment. *Journal of Personality Assessment*, 85, 219–237. doi:10.1207/s15327752jpa8502\_16
- \*Soilevuo Grønnerød, J., & Grønnerød, C. (2010). Are large drawings signs of psychological expansion or effects of drawing skills? A critical evaluation of Wartegg drawing size categories in a Finnish sample. *Scandinavian Journal of Psychology*, 51, 63–67. doi:10.1111/j.1467-9450.2009.00729.x
- Soilevuo Grønnerød, J., & Grønnerød, C. (2011). *The Wartegg bibliography* [An electronic collection of references]. Retrieved September 20, 2011, from <http://www.citeulike.org/profile/wartegginfo>
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. (1977). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists Press.
- Sugiura, K., Hara, S., Suzuki, Y., Takeuchi, M., & Kakudate, N. (2005). Characteristics of junior high school students and high school students on projective drawing technique test battery: The first report. *Bulletin of Liberal Arts & Sciences: Nippon Medical School*, 35, 37–61.
- Sugiura, K., & Takanashi, R. (2001). A report on projective drawing technique test battery. *Bulletin of Liberal Arts & Sciences: Nippon Medical School*, 31, 11–31.
- Sugiura, K., & Yagi, S. (2002). The significance of a projective drawing technique test battery for parents with children who refuse to attend school: Changes in the mothers discussed with reference to the projective drawing technique test battery. *Bulletin of Liberal Arts & Sciences: Nippon Medical School*, 32, 63–93.
- \*Takala, M. (1953). Studies of psychomotor personality tests I. *Annales Academiae Scientiarum Fennicae, Serie B*, 81, 1–130. Helsinki, Finland: Suomalainen Tiedakatemia.



- Takala, M. (1957). *Analyttinen arvostelumenetelmä Warteggin piirtämistestistä varten* [An analytic scoring system of the Wartegg Drawing Completion Test]. (No. 12). Jyväskylä, Finland: University of Jyväskylä, College of Education, Department of Psychology.
- \*Takala, M. (1964). Studies of the Wartegg drawing completion test: Studies of psychomotor personality tests II. *Annales Academiae Scientiarum Fennicae, Serie B, 131*, 1–112. Helsinki, Finland: Suomalainen Tiedekatemia.
- Takala, M., & Hakkarainen, M. (1953). Über Faktorenstruktur und Validität des Wartegg-Zeichen-testes [Factor analysis and validity of the Wartegg Drawing test]. *Annales Academiae Scientiarum Fennicae, Serie B, 81*, 1–95.
- \*Takala, M., & Rantanen, A. (1964). Psychomotor expression and personality study II. *Scandinavian Journal of Psychology, 5*, 71–79. doi:10.1111/j.1467-9450.1964.tb01411.x
- Takayanagi, K. (2008). Psychological and physical change at the green area in the middle of the urban square. *Japanese Journal of Complementary and Alternative Medicine, 5*, 145–152. doi:10.1625/jcam.5.145
- \*Tamminen, S., & Lindeman, M. (2000). Wartegg—luotettava persoonallisuustesti vai maagista ajattelua? [The Wartegg—A valid personality test of magical thinking?]. *Psykologia, 35*, 325–331.
- \*Teiramaa, E. (1977). *Psychosocial factors in the onset and course of asthma: A clinical study on 100 patients* (Acta Universitatis Ouluensis, Series D, Medica No. 14, Psychiatrica No. 4). Oulu, Finland: University of Oulu.
- \*Teiramaa, E. (1978a). Psychic disturbances and duration of asthma. *Journal of Psychosomatic Research, 22*, 127–132. doi:10.1016/0022-3999(78)90039-9
- \*Teiramaa, E. (1978b). Psychic disturbances and severity of asthma. *Journal of Psychosomatic Research, 22*, 401–408. doi:10.1016/0022-3999(78)90062-4
- \*Teiramaa, E. (1979a). Asthma, psychic disturbances and family history of atopic disorders. *Journal of Psychosomatic Research, 23*, 209–217. doi:10.1016/0022-3999(79)90006-0
- \*Teiramaa, E. (1979b). Psychic factors and the inception of asthma. *Journal of Psychosomatic Research, 23*, 253–262. doi:10.1016/0022-3999(79)90027-8
- \*Teiramaa, E. (1979c). Psychosocial and psychic factors and age at onset of asthma. *Journal of Psychosomatic Research, 23*, 27–37. doi:10.1016/0022-3999(79)90068-0
- \*Teiramaa, E. (1981). Psychosocial factors, personality and acute-insidious asthma. *Journal of Psychosomatic Research, 25*, 43–49. doi:10.1016/0022-3999(81)90082-9
- Testilautakunta. (2008). *Projektiivisten menetelmien käyttökelpoisuudesta* [On the usability of projective tests]. Retrieved February 18, 2008, from [http://www.testilautakunta.fi/Projektiiviset\\_menetelmat.pdf](http://www.testilautakunta.fi/Projektiiviset_menetelmat.pdf)
- \*Togliatti, M. M., Masci, M., Ciardi, A., Gargano, T., Lombardi, R., Micci, A., . . . Gallo, P. (2003). Valutazione della personalità e della modifica dei comportamenti a rischio in un campione di studenti: Risultati e riflessioni [Personality assessment and modification of the behaviors at risk in a students' sample: Results and reflections]. *Notiziario dell'Istituto Superiore di Sanità, 18*, 14–16.
- \*Venturino, G., Sciaudone, G., & Scotto di Tella, A. (1994). Abitudini alcoliche e profilo psicologico di un campione di donne gravide a confronto con un gruppo di controllo [Alcohol consumption and personality profile in a group of pregnant women]. *Medicina Psicomatica, 39*, 101–119.
- Vetter, A. (1952). *Der Deutungstest (Auffassungstest) Wartegg-Vetter: Ein diagnostisches Hilfsmittel für die psychologische Beratung* [The Interpretation Test of Wartegg-Vetter: A diagnostic aid for psychological assessment]. Stuttgart, Germany: Testverlag S. Wolf.
- Wartegg, E. (1939). *Gestaltung und Charakter: Ausdrucksdeutung zeichnerischer Gestaltung und Entwurf einer charakterologischen Typologie* [Form and character: Interpretation of expression from drawings and an outline of a characterological typology]. Leipzig, Germany: Verlag von Johann Ambrosius Barth.
- Wartegg, E. (1953). *Schichtdiagnostik: Der Zeichentest (WZT): Einführung in die experimentelle Graphoskopie* [Layered diagnosis: The Drawing Test (WZT): An introduction to experimental graphoscopy]. Goettingen, Germany: Verlag fuer Psychologie.
- Wartegg, E. (1954). Der Zeichentest (WZT): Einführung in die graphoskopische Schichtdiagnostik. In E. Stern (Ed.), *Handbuch der klinischen Psychologie: Band 1. Die Tests in der klinischen Psychologie: 2. Halbband* [Handbook of clinical psychology: Vol. 1. Tests in clinical psychology: 2nd half volume] (pp. 520–587). Stuttgart, Germany: Rascher Verlag.
- Wartegg, E. (1959). *Reattivo di disegno per la diagnostica degli strati della personalità: Manuale: Adattamento italiano* [The use of drawings in layered diagnosis of personality: Manual: Italian adaptation]. Firenze, Italy: Organizzazioni Speciali.
- Wartegg, E. (1960). *Wartegg Test de personalidad grafico-proyectivo* [The Wartegg: A graphic-projective test]. Bogota, Colombia: PSEA.
- Wartegg, E. (1972). Il reattivo di disegno (WZT): Introduzione, note e casistica italiana [The Wartegg Drawing Completion Test: Introduction, note and Italian case studies]. Firenze, Italy: Organizzazioni Speciali.
- Wartegg, E., D'Alfonso, P. G., & Biedma, C. J. (1965). *Test de Wartegg-Biedma* [The Wartegg-Biedma Test]. Neuchatel, France: Delachaux & Niestlé.
- \*Wass, T., & Mattlar, C.-E. (2000). *Warteggs teckningstest: Manual* [The Wartegg Drawing Test: A manual]. Stockholm, Sweden: Psykologiförlaget AB.
- Wassing, H. E. (1974). Van Krevelen's modification of the drawing-completion test versus Wartegg's original version: A comparative examination of results. *Acta Paedopsychiatrica, 40*, 122–136.

(Appendix follows)



## Appendix

## Meta-Analysis Moderator Coding Criteria

Table A1

Variable	Code	Description
<i>Study codes</i>		
Publication year		
Country		Country of first author (journal article, dissertation) or publication country (books). Germany used for 1933–1948 and 1990 to present; otherwise DDR or BDR from affiliation location.
Language		Language used in publication
Publication type	3	Doctoral thesis
	4	Monograph, book, book chapter
	5	Journal article, full journal, serial
Scoring system	0	Not Reported
	1	Wartegg and others (Avé-Lallemant, 1978; Lossen & Schott, 1952; Renner, 1953; Vetter, 1952; Wartegg, 1939, 1953)
	2	South American systems (Biedma & d'Alfonso, 1960; Freitas, 1993)
	3	Gardziella and others (Gardziella, 1985; Puonti, 2005; Wass & Mattlar, 2000)
	4	Crisi (1998)
	5	Kinget (1952)
	6	Takala (1957)
	7	Kukkonen (1962a, 1962b)
	8	Several scoring systems
Reported result	0	WZT mentioned as one of the methods used, but no specific WZT results reported
	1	WZT results presented but not as codable results
	2	Codable WZT results
<i>Sample codes</i>		
Design	0	Single sample
	1	Comparative samples, no control group
	2	Comparative samples with control group
Other methods		Free response methods
		Interview
		Self-report
		Ability/cognitive
		Medical
		Other
Scorer blinding	0	Not Reported/No Blinding
	1	Partial Blinding
	2	Full Blinding
Reliability check	0	Not Reported/No Check
	1	Partial check (inappropriate method used)
	2	Full check
<i>Result codes</i>		
Subject population	0	Non-Patients, General
	1	Non-Patients, Selection
	2	Somatic Patients
	3	Outpatients
	4	Inpatients
	5	Others/Mixed
Subject age		Mean and standard deviation, range, general group (e.g., adults), adults coded as 38 based on average of other studies reporting adult age
Focused statistics	0	Not focused: $F$ test or $\chi^2$ test with $df < 1$
	1	Focused: all other statistics
Sign	-1	Result in the opposite direction of a clearly stated hypothesis
	0	No clear hypothesis related to the result
	1	Result in the expected direction of a clearly stated hypothesis

(Appendix continues)

Table A1 (continued)

Variable	Code	Description
Criteria	0	Self-Report
	1	Free Response
	2	Observer
	3	Diagnosis (somatic and psychiatric)
	4	Other (social class, age, handwriting, family history)

*Note.* Additional codes were initially used but did not pertain to the meta-analysis studies. DDR = former East Germany; BDR = former West Germany; WZT = Wartegg Zeichen Test.

Received December 20, 2010  
Revision received June 9, 2011  
Accepted June 21, 2011 ■