

Inter Scorer Reliability of Hand Test

Fayyaz Ahmed Anjum & Iffat Batool
Department of Psychology
GC University Lahore

The present study was carried out to establish the psychometric properties (inter scorer reliability) of the Modified Hand Test. The Modified Hand Test consists of 14 stimuli including four new stimuli, which were adapted through a study to determine cultural input in accordance with the original test construction procedure. This study was carried out on a sample of 50 participants who were diverse in terms of age, educational level and economic backgrounds. The sample was selected randomly out of the larger sample of 500 participants who were recruited for the establishment of norms of the Hand Test in Pakistan. Three scorers, including the researcher, scored the responses independently in accordance with guidelines provided in the manual of the Hand Test. Agreement was considered as very high between pair of scorers. A high level of agreement was found between scorers one and two of 88%, scorers two and three of 86%, and scorers one and three of 87%. Comparison of agreement between pair of examiner were also made for the Modified Hand Test which consisted of 14 stimuli with the Original Hand Test comprising of ten stimuli, similar results were found.

The Hand Test was developed by Wagner in 1962 and new norms for children and adolescents were introduced in 1983. The Hand Test has gained the reputation of a reputable projective and diagnostic measure and became popular among clinicians and researchers because of its capacity to measure various aspects of human personality and behavior (Naeem, 1979). Despite being a projective test, it offers objectivity in terms of well-defined quantitative scoring systems (Naeem, 1979; Wagner, 1983).

The Hand Test has four major domains: Interpersonal, Enviromental, Maladjustive and Psychotic. The first domain of Interpersonal domain encompasses affection, dependence, communication, exhibition, direction, and aggression. Affectionate responses involve an interchange or bestowment of pleasure, affection or friendly feelings for instance "waving to a friend", "a greeting, shaking hands." Dependence responses involve an expression of dependence or need for succorance from another person. For example, "hitch hiker", "thumbing a ride", "seeking help", "praying for help and begging." Communication responses involve a presentation or exchange of information. For example, "discussion", "explaining something to someone", "is conveying message". Exhibition: responses displaying or exhibiting one to obtain approval from others or to seek attention of others. For example, "showing off his muscles", "nail polish", "engagement ring." Direction: responses involve influencing the activities or domineering or directing others. For example, "policeman saying stop", "giving orders to someone", "guiding or directing a child by a parent not to do this." Aggression: responses involve giving of pain, hostility or aggression. For example, "trying to scare someone", "insulting or inflicting pain", "slapping", "showing threat" (Sergio, 2010; Shnider, 1973; Wagner, 1983).

The second domain of Environmental responses reflect a person interaction with environment, its sub categories are action, acquisition and passive whereby action and acquisition have positive connotation in terms of demonstration of performance and

the action is yet to be completed, whereas responses are reflected in the context of environment. Passive category represents unproductive aspects or lack of energy of individual. These two main categories (interpersonal, environmental) represent normal protocol of personality (Wagner, 1983).

The third domain of maladjustive category represents generally neurotic protocol (Wagner, 1983) and consists of tension, crippled and fear responses whereby tension represents anxiety and inadequacy and crippled indicates inferiorities and fear reflects avoidance pattern of a person from danger perceived in the outer environment (Sergio, 2010; Wagner, 1983).

The fourth main domain of withdrawal category represents psychotic protocol Wagner (1983) and consists of description, fail and bizarre whereby in description responses people tend to describe the stimulus position instead of following instructions to respond what the hands are doing (Wagner, 1983). In fail participants demonstrate their inability to perceive any response probably due to their perception of threatening environment. Bizarre responses represent inability of any real and meaningful contact with environment and people around, which is depicted in terms of hallucination and delusional responses (Wagner, 1983).

Only few researches have been carried out in the west to determine inter scorer reliability of Original Hand Test since its inception in 1962.

The first study was conducted by Wagner, Maloney and Wilson (1979) on normal female participants with mean age of 20.6 years and SD =1.79 and the agreement between two examiners was 89.34 %. Spearman correlation was also calculated between quantitative and combined scoring categories which ranged from .71 to 1 (perfect agreement).

Another investigation was carried out by Wendler and Zachary (1983) on a sample of 65 protocols of normal, mentally retarded and maniac depressive participants. Kappa coefficient for sub and main scoring categories ranged from .45 to .1 with overall .69 reliability coefficient where communication produced lowest kappa i.e. .45 and highest kappa was found in failure and exhibition as 1 each whereas for the combine categories it remained .92 for interpersonal, .75 for environmental, .83 for maladjustive and .71 for withdrawal. The overall kappa for the combined category was as .81. The second part of the study consisted of agreement between two scorers which remained as 72% for the 15 sub categories and

87% for the combined scoring categories. Low agreement was found where responses were given in fewer numbers. For example, exhibition percentage of agreement for combined scoring categories remained from 78 to 96%. Disagreement was found in combined categories rather than sub categories.

High percentage of agreement was also found among three scorers on 100 Hand Test protocols while determining internal consistency reliability cited in Hand Test Manual (Wagner, 1983, p. 67). Three scorers were graduate and were provided Hand Test manual as a guide. The criteria for agreement was determined as a perfect agreement on the 15 quantitative scoring categories. Omissions and mismatched were considered as disagreement. The percentage of agreement remained 80 % between scorer one and two and 78% between scorer one and three whereas it remained 83% between scorer two and three. Most of the disagreement was found between sub categories. For example, acquisition and action instead of combined scoring categories.

Only one investigation had also been previously carried out in Pakistan by Zehra (1989) on 20 participants who were randomly selected out of a larger sample of 113 participants, whereas the mean age of participants remained 15.05 years with $SD=1.10$. The responses of these participants were scored by three examiners including the researcher, a PhD scholar and a Clinical Psychologist who had a wide experience of teaching, practice and research especially on the Hand Test. The agreement between the scorers one and two was 79.5 %, between scorer one and three was 70.51 % and between scorers two and three was 68 %. The researcher recommended that inter scorer reliability may be conducted on a divergent sample.

Rationale of the present study

The aim of the study was to examine the psychometric properties i.e. interscorer reliability of The Modified Hand having cultural reference.

Method

Sample

In total, 50 participants were randomly selected out of 500 participants from a sample of larger study for development of Pakistani norms for the Hand Test. Purposive sampling technique was used to recruit the participants. The mean age of the sample was as 34.42 with $SD = 17.17$. Seventeen participants (nine males and eight females) were adolescents studying in class ninth and tenth while the other 33 participants were graduate and post graduate. 14 belonged to middle adulthood whereas 17 to later adulthood. They belonged different educational institutions representing different economic classes from Islamabad, Pakistan.

Instrument

Modified Hand Test

The Modified Hand Test consists of fourteen stimuli. Ten from the Original Hand Test and four new stimuli were adapted in accordance with test construction procedure (Riaz, 2008). Ten stimuli were opted by a panel of experts and were administered on a sample of 100. These stimuli were selected on the criteria of a stimulus ability to generate cognitive and affective responses in

more scoring categories and varied range of responses in each scoring category. Stimuli were increased to enhance reliability of the instrument and to ensure cultural input from participants of this region.

Procedure

The test was administered individually by the researcher and was scored in accordance with standardized procedures given in the manual of the Hand Test (Wagner, 1983). The test consisted of fourteen stimuli. In this modified version, four additional stimuli were adapted in accordance with test construction procedure (Kline, 1986; Riaz, 2008) and were added after stimulus nine. These stimuli were added in sequence of stimulus ten, eleven, twelve and thirteen, the last tenth stimulus in the Original Hand Test which is an invitation to imagination was now placed again at number fourteenth. Stimulus were adapted with agreement from the publisher Western Psychological Services. The responses were scored independently by three scorers and agreement was considered as perfect, omissions and mismatch were considered as disagreement, percentages of agreement between pair of examiners were computed i.e., between scorer one and scorer two, scorer one and scorer three and scorer two and scorer three on each stimulus.

The three scorers were experienced. The first scorer is a researcher a PhD scholar who used the Hand Test in M.S project. The second scorer who had wide experience of selection of personnel in armed forces of Pakistan with M.Phil qualification and who had experience of using this instrument in Masters level research. The third scorer was a student of MS Clinical psychology and was trained in scoring of the Hand Test, was aided with example of responses of different scoring categories given in the manual. Agreement of three scorers have been computed for Modified Hand Test consisting of fourteen stimuli as well as for the Original Hand Test consisting of ten stimuli.

Results and Discussion

To determine inter scorer reliability of the Modified Hand Test in Pakistan. Results have been computed (see Table 1) and they demonstrated a high level of agreement among the three scorers.

Inter-scorer agreement was as 88 % between scorer one and two, and was 87% between scorer two and three, and was 86% between scorer one and three. Perfect agreement was found between scorer one and two on stimulus thirteen which also depicted stimulus constancy (ability to elicit response in specific category) of this stimulus for dependence variable whereas the lowest agreement was found between scorer one and three on stimulus three, this also suggests that the experienced scorer had scored the responses as communication or direction whereas the novice scorer scored it as action. This result is in line with the findings of investigation carried out by Wagner, Maloney, and Wilson (1981) that the difference in scoring was within scoring categories instead of between main scoring categories. Moreover, the finding of this study is also supported by the results of study carried out by Zehra (1989) in Pakistan as results of her study suggested that the difference in scoring of responses were found between main scoring categories. In the present study, the disagreement between scorer one and three and scorer two and three are highlighted because the responses were scored between scorer one and two were within the scoring categories and disagreement between scorer one and three and scorer two and three were within the main scoring categories. It

Table 1
Percentage Agreement amongst Three Scorers on Various Stimulus of Modified Hand Test

Stimulus	Scorers 1 & 2			Scorers 1 & 3			Scorers 2 & 3		
	A	D	% of A	A	D	% of A	A	D	% of A
I	43	4	86%	44	3	88%	46	1	92%
II	41	6	82%	46	1	92%	42	5	84%
III	41	6	82%	31	16	62%	37	10	74%
IV	46	1	92%	45	2	90%	46	1	92%
V	45	2	90%	43	4	86%	45	2	90%
VI	43	4	86%	45	2	90%	45	2	90%
VII	46	1	92%	41	6	82%	42	5	84%
VIII	44	3	88%	46	1	92%	45	2	90%
IX	41	6	82%	43	4	86%	45	2	90%
X	45	2	90%	40	7	80%	42	5	84%
XI	41	6	82%	44	3	88%	44	3	88%
XII	46	1	92%	45	2	90%	41	2	82%
XIII	50	-	100%	45	2	90%	45	2	90%
XIV	48	1	96%	45	2	90%	46	1	92%
	62								
Total	0	43	88%	645	55	86%	657	43	87%

A= Agreement D= Disagreement

is to highlight that comparison of results based on fourteen stimuli of the Modified Hand Test with the Original Hand Test provided almost similar results which were scored by scorer (1 & 2= 87%, scorer 1 & 3= 85% and scorer 2 & 3= 87%).

Disagreement between scorer one and three was found on stimulus three in both studies carried out in Pakistan. The present study also suggested that inter scorer agreement is dependent upon three factors. First, the experience of scorers with the instrument. Second, the type of sample whether heterogeneous or homogenous and number of sample. Third, the specific or generic pull of stimulus in the present study, perfect agreement was found on stimulus thirteen which proved to have specific pull for dependence scoring category.

High inter scorer reliability of Modified Hand test reflected authenticity of procedure implied for selection and adaptation of new stimuli, enhanced stimuli have direct relationship with increased reliability of Modified Hand Test. The test can be confidently utilized in assessment of personality and its domains. Further researches would bring fruitful results.

In conclusion, inter scorer reliability has been carried out on a limited sample of 50 participants. Future researches can be carried out on a larger and diversified sample (Comprising of normal, neurotic and psychotic sample) for generalization of the results. Inter-scorer agreement among three scorers are in support of well-defined quantitative scoring procedures of the Original Hand Test and for the Modified Hand Test.

References

- Anastasi, A. (1997). *Psychology, psychologists, and psychological testing (1967)*. In J. M. Notterman, & J. M. Notterman (Eds.), *The evolution of psychology: Fifty years of the American Psychologist* (pp. 373-390). Washington, DC, US: American Psychological Association. doi:10.1037/10254-027
- Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. London: Methuen.
- Naeem, P. (1979). *Psychological profile of professional and nonprofessional criminals* (Unpublished Ph.D Dissertation). National Institute of Psychology, Quaid-i-Azam University, Islamabad, Pakistan.
- Panek, P. E., & Hayslip, B. (1980). Construct validation of the Hand Test withdrawal score on institutionalized older adults. *Perceptual and Motor Skills*, 51(2), 595-598.
- Riaz, M. N. (2008). *Test construction, development and standardization of psychological tests in pakistan*. Pakistan: Higher Education Commission,
- Sergio, J. A. (2010). *Differences among abused and non-abused younger and older adults as measured by the Hand Test*. *Dissertation Abstracts International*, 72, 1196. University of North Texas.
- Waehler, C. A., Rasch, M. A., Sivec, H. J., & Hilsenroth, M. J. (1992). Establishing a placement index for behaviorally disturbed children using the Hand Test. *Journal of Personality Assessment*, 58(3), 537-547.
- Wagner, E. E. (1962). *The Hand Test: Manual Revised*. (1962). PLACE: Western Psychological Services.
- Wagner, E. E. (1983). *The Hand Test: Manual Revised*. (1983). PLACE: Western Psychological Services.
- Wagner, E. E., Maloney, P., Wilson, D. G. (1981). Split half and test retest reliability for pathological sample. *Journal of Clinical Psychology*, 37, 589-592
- Wendler, C. L. W., & Zachary, R. A. (1983). *Reliability of scoring categories on a projective test*. Paper presented at the meeting of the Western Psychological Association.

Zehra, F. (1989). Inter scorer reliability of Hand Test in Pakistan,
Pakistan Journal of Psychological Research 4 (1-2), 43-46.

Received: 9th November, 2016
Revisions Received: 1st May, 2018