INTRODUCTION

After examinees complete a psychological assessment in a psychiatric setting, clinicians judge whether the examinee is making reasonable efforts in performing the psychological test, is reporting exaggerated symptoms, or conversely, under-reporting their symptoms. Testifying the validity of the psychological assessment is very important for both the evaluators and the clinicians who receive the report. Specifically, the examinee may show a poor effort in the psychological test by over- or under-reporting their symptoms for various reasons, including legal issues, financial gain, stigmas, or refusal of treatment; clinicians must be aware of this to obtain valid data. However, it is not easy for the clinicians to determine whether the examinee has made best efforts in the psychological test based on performance-based measures, self-reported measures, and clinical judgments. Thus, ensuring the reliability of the psychological test is very difficult.

Generally, self-reported measures have validity scales embedded in the tests which could help clinicians determine the over- or under-reporting of symptoms reported by examinees. The F scale in the Minnesota Multiphasic Personality Inventory-2 (MMPI-2), one of the most well-known self-reported measures, is a scale that can effectively detect examinees who exaggerate psychopathology in the tests. The L or K scale in the MMPI-2 can effectively detect under-reporting. In another self-reported measure known as the Personality Assessment Inventory (PAI), the negative impression (NIM) scale measures distorted responses to provide a negative impression, while positive impression (PIM) measures the attempts...
to impress to provide a positive impression.

In contrast, performance-based measures often do not include the validity scales that are included in the self-reported measures. Determining validity in performance-based measures entirely depends on the examinee’s performance efforts. In this case, it may be difficult to judge poor performance or malingering unless clinicians have an objective basis for judging the examinee’s performance validity. The most commonly used performance-based measures in psychiatry can be intelligence tests, and the most recent Wechsler Adult Intelligence Scale-IV (WAIS-IV) is one of the most widely used adult intelligence tests. Similarly, it may be difficult for the clinician to judge whether or not proper efforts have been made by the examinee in WAIS-IV. In this study, the Korean-Wechsler Adult Intelligence Scale-IV (K-WAIS-IV) was used to verify the examinee’s performance validity. The most common performance-based measures in psychiatry can be intelligence tests, and the most recent Wechsler Adult Intelligence Scale-IV (WAIS-IV) is one of the most widely used adult intelligence tests. In this study, the Korean-Wechsler Adult Intelligence Scale-IV (K-WAIS-IV) was used to verify the performance validity of examinees and to suggest the indicator of performance validity so that it can be easily utilized in actual psychiatric settings. First, a brief review of the tests that allow the clinician to determine the validity of performances or symptoms will be followed.

There are two main types of test that can identify poor performance and response bias. One is Symptom Validity Tests (SVTs) that can detect over- and under-reporting symptoms, and the other is Performance Validity Tests (PVTs) that can detect poor performance. PVTs detect poor performance and are mainly used in neuropsychological and intelligence tests. PVTs were initially used in forensic settings with a high rate of malingering patients, and were often helpful in determining malingering severity. Next, SVTs detect over-reporting or false reports of symptoms and can be used in self-reported measures or structured interviews. In typical self-reported measures such as MMPI-2, the clinician can detect falsified symptoms or over-reporting by examinees through the validity scales.

PVTs can be categorized into two types. The embedded type that utilize a part or a subtest that belongs to an individual test. The stand-alone type is a completely individual test. Embedded types include the Digit Span of the Wechsler intelligence scales, the Rey-Osterrieth Complex Figure Test of the Rey Memory Test, and so forth. Stand-alone types include the California Verbal Learning Test, the Test of Memory Malingering, the Word Memory Test, and so forth. Particularly, embedded-type PVTs do not require additional time and costs, and they can be easily and quickly used in a clinical setting.

One of the embedded-type PVTs called the Reliable Digit Span (RDS) is a PVT that uses the Digit Span subtest score on the Wechsler Adult Intelligence Scale (WAIS). RDS is one of the most studied PVTs and is one of the most commonly used PVTs in various fields. Greiffenstein et al. first used the Digit Span subtest on the Wechsler adult intelligence scale to develop RDS. They considered the inconsistent memory test results in head trauma patients as a secondary gain. Based on this, they derived RDS from the Wechsler intelligence scale. RDS is calculated by using the Longest Digit Span Forward (LDSF) and Longest Digit Span Backward (LDSB) on the WAIS.

In previous studies, researchers attempted to find appropriate cut-off scores for RDS in various groups. Greiffenstein et al. regarded an RDS score of 7 pts as the appropriate cut-off score. A subsequent review study on patients with Traumatic Brain Injury (TBI), intellectual disabilities, adult Attention Deficit Hyperactivity Disorder (ADHD), and chronic pain found that an RDS score of 7 pts was the appropriate cut-off score for the group of TBI, adult ADHD, and chronic pain patients. However, the study found that an RDS score of 6 pts was the appropriate cut-off score for the groups with severe memory disorders, intellectual disabilities, borderline intelligence, and those whose mother tongue is not English. Another previous study with a psychosis group (consisting of patients with the schizophrenia spectrum disorders and other psychotic disorders) and a non-psychosis group (consisting of patients with major depressive disorder, generalized anxiety disorder, and post-traumatic stress disorder) found that an RDS score of 7 pts was a more appropriate cut-off score. In addition, previous studies on mixed clinical patients (depressive disorders, pain disorders, seizure disorders, and etc.) suggested that an RDS score of 6 pts was the appropriate cut-off score. In contrast, previous studies with persistent post-concussive patients, neurocognitive dysfunction patients, and university students showed that an RDS score of 7 pts was the appropriate cut-off score.

Some studies have suggested an RDS score of 5 and 3 pts. One with cognitive impairment patients found that an RDS score of 5 pts was the appropriate cut-off score. The other one with participants with low IQs of 70 or below suggested that an RDS score of 3 pts was the appropriate cut-off score.

Taken all together, the results of previous studies seem to consider RDS scores of 6 or 7 as the appropriate cut-off score. However, in the patient group with organic deficits such as TBI or intellectual disabilities, RDS scores of 3 or 5 pts can be the appropriate cut-off score. Therefore, RDS cut-off scores are expected to differ from different patient groups. In addition, Lippa reported that few RDS studies have focused on non-English speakers, especially Asian patients. Therefore, the RDS cut-off score may be different from Asian patients than from English-speakers because of cultural differences.

Meanwhile, the studies on RDS have rarely studied in South Korea. The review article from Jeong and Choi briefly mentioned RDS as one of the WAIS-based indicators for detecting malingering. Similar to RDS, Moon and Hwang suggested the
discriminant function formula using WAIS in order to detect malingering in patients with mild TBI. The formula included WAIS subtests such as Digit Span, Vocabulary, Visual Puzzles, and Coding subtests. Moreover, among the subtests included in the discriminant function formula, Moon and Hwang noted that the Digit Span subtest was the best discriminator for malingering. Moon and Hwang suggested the more studies on malingering in South Korea are needed and that fundamental epidemiologic studies including base rate by group should be preceded first.

The aim of this study is to suggest the cumulative frequency and base rate of RDS and to provide the RDS cut-off score for epidemiologic data and malingering detection in the psychiatric field. The base rate here refers to the prevalence or the cumulative frequency of the observed score. A low RDS base rate refers to a low prevalence or cumulative frequency of the RDS score. Typically, being statistically significant based on p values and having a low prevalence as per the base rate are independent of each other. A specific score can be low with statistical significance, but the prevalence based on the base rate or the cumulative frequency can be quite common. A previous study examining the RDS base rate showed that an RDS score corresponding to 10% of the base rate can be regarded as the cut-off score. Therefore, an RDS score corresponding to 10% of the base rate was considered as cut-off score in this study.

METHODS

Participants
Participants included in- and out- patients who visited the psychiatry department from December 2016 to November 2018. They were divided into 4-group by psychiatrists based on the Diagnostic and Statistical Manual of Mental Disorder Fifth Edition (DSM-5): 1) patients visiting for military service issues (military service group, n=61); 2) patients visiting for disability and compensation assessment after being diagnosed with major or mild neurocognitive disorders due to TBIs (TBI group, n=57); 3) patients with psychosis including the schizophrenia spectrum and other psychotic disorders and bipolar disorder (psychosis group, n=47); 4) patients with depressive and anxiety disorders (neurosis group, n=96). The study was conducted after obtaining the approval of the Institutional Review Board of the Wonkwang University Hospital (IRB No. WKUH 2019-07-015).

Measures

Reliable Digit Span of K-WAIS-IV
The Korean version of WAIS-IV (K-WAIS-IV) by Hwang et al. was used in this study. K-WAIS-IV consists of ten core subtests and five supplemental subtests. The Digit Span subtest from ten core subtests was used to calculate RDS from K-WAIS-IV. The Digit Span subtest is a core subtest of the Working Memory Index (WMI) and consists of the tasks Digit Span Forward (DSF), Digit Span Backward (DSB), and Digit Span Sequencing (DSS). These three tasks each run the same Digit Span of numbers twice with increasing difficulty as the task progresses due to the increasing Digit Span of the numbers. In the Digit Span subtest, the clinician can calculate the raw scores and process scores for DSF, DSB, and DSS.

RDS can be calculated as follows. Among the process scores in the Digit Span subtest, RDS is calculated by adding the scores from the Longest Digit Span Forward (LDSF) and the Longest Digit Span Backward (LDSB). However, when calculating RDS, it is calculated with the longest Digit Span score where two questions are answered correctly in the trial with the same Digit Span. Caution is required as it is different from the method of calculating the longest Digit Span score that is

<table>
<thead>
<tr>
<th>Table 1. Example of Reliable Digit Span calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Digit Span forward</strong></td>
</tr>
<tr>
<td><strong>Trial</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td><strong>Reliable Digit Span score=5pts</strong></td>
</tr>
</tbody>
</table>

*Reliable Digit Span score: Longest Digit Span forward score + Longest Digit Span backward score (ex. 3 span + 2 span → Reliable Digit Span score 5 pts.), Longest Digit Span score where two questions must be answered correctly in the trial with the same Digit Span
required in the K-WAIS-IV administration and scoring manual. The illustrative example is shown at Table 1.

**Data analysis**
Frequency and descriptive analysis were conducted to examine the demographic characteristics. Age and level of education were presented as average and standard deviations, while gender was presented as the number of cases and a percentage. The indexes and the scaled scores by group were presented as average and standard deviations through frequency analysis. Next, frequency analysis was conducted in order to examine the cumulative percentage and base rate of RDS, and the results were presented as a cumulative frequency and base rate (cumulative percentage). According to a previous study, an RDS score corresponding to 10% of the base rate through frequency analysis was considered as the cut-off score in this study. In order to examine the difference in RDS scores among the groups, ANOVA was conducted. Moreover, η², which represents the effect size in ANOVA, was calculated; η² values of 0.01 to 0.06 represent a “small effect,” 0.06 to 0.14 represent a “medium effect,” and 0.14 and above represent a “large effect.” This study used SPSS 21.00 (IBM Corp., Armonk, NY, USA) for statistical analysis.

**RESULTS**

**Demographic data**
The demographic data of this study are as follows (Table 2). The average age by group was 21.22, 57.05, 32.10, and 34.29 years old for the military service, TBI, psychosis, and neurosis groups, respectively. Next, the number of cases for each gender showed 61 males in the military service group which corresponded to 100%. TBI group showed 48 males, corresponding to 84.21%, while the number of females was 9, corresponding to 15.79%. In the psychosis group, there were 21 females corresponding to 44.68%, and 26 females corresponding to 55.32%. The neurosis group had 36 females corresponding to 37.50%, and 60 females corresponding to 62.50%.

**K-WAIS-IV scores by group**
The FSIQ score of the military service group was 83.21 pts. The FSIQ scores of the TBI, psychosis, and neurosis groups were 67.96 pts, 77.38 pts, and 83.51 pts, respectively. The indexes of K-WAIS-IV and the scaled scores of subtest by group are presented in Table 3.

**Base rate of RDS**
The RDS score corresponding to 10% of the base rate for each group was confirmed (Table 4). It was 7 pts for the military service group, 6 pts for the neurosis and psychosis groups, and 3 pts for the TBI group.

<table>
<thead>
<tr>
<th>Table 2. Demographic characteristics by group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (year), mean±SD</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Age, N=61</td>
</tr>
<tr>
<td>Sex, %</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Education (year), mean±SD</td>
</tr>
</tbody>
</table>

Military service: group of patients visiting for military service issues. TBI: traumatic brain injury.
Table 5 shows the difference in RDS scores among the military service, TBI, neurosis, and psychosis groups. The average RDS was 10.65 pts (SD=2.58 pts) for the military service group, 5.92 pts (SD=2.61 pts) for the TBI group, 9.02 pts (SD=2.52 pts) for the psychosis group, and 9.59 pts (SD=2.63 pts) for the neurosis group. In the test for differences in the averages, a significant difference among the groups was observed (F=36.35, p<0.001). In addition, η² was found to be 0.30, corresponding to a "large effect." When the difference among the groups was examined in detail using post-hoc analysis, the RDS score for the TBI group was significantly lower than that of other groups. The RDS score for the military service group was significantly higher than that of the psychosis group. The RDS score for the neurosis group did not show a significant difference from that of the military service or the psychosis group but was still significantly higher than that of the TBI group.

**DISCUSSION**

RDS in WAIS is one of the representative PVTs that can assess poor performance by examinees. This study attempted to provide fundamental epidemiological data for the clinical field by calculating the cumulative frequency and base rate of RDS using K-WAIS-IV. Thus, the present study analyzed the RDS base rate by group (military service, TBI, psychosis, and neurosis) and introduced the RDS cut-off score corresponding to 10% of the base rate. The study also testified the difference in RDS scores among the groups.

As with previous studies, this study regarded 10% of the base rate as cut-off.6,41 The RDS scores corresponding to 10% of the base rate for each group were 7 pts for the military service group, 3 pts for the TBI group, psychosis groups, and 6 pts for the neurosis. The each RDS scores can be regarded as the cut-off scores for the each group.6,41

The results of this study were compared with that of previous studies that suggested RDS cut-off scores by receiver operating characteristic analysis (ROC). The previous studies on persistent post-concussive patients, neurocognitive dysfunction patients, and university students suggested 7pts of RDS cut-off score, which is the same cut-off score for the military service group of this study.32-34 Furthermore, 6 pts of RDS cut-off score for the neurosis and psychosis groups in this study was coincided with the results of the previous study on mixed clinical patients and chronic pain patients.30,31

On the other hand, for the TBI group in this study, the RDS cut-off score was 3 pts, which was coincided with the results of the previous study on patients with low IQs of below 70.36 In the present study, the RDS score of the TBI group was relatively lower than that of the military service, psychosis, and
neurosis groups. The reason could be the effects of brain damage so that the participants with TBI had more difficulties with performing cognitive tasks than the military service group.36

In summary, the RDS cut-off score was 7, 6, and 3 pts for the military service, neurosis and psychosis, and TBI groups, respectively. If the RDS cut-off scores or lower scores were shown in practice, the clinician might suspect intentional poor performance or malingering attitude. Practically, if the examinees with the military service issues show RDS 7 or less pts, high chance of malingering could be suspected.

However, the clinician should not utilize the RDS cut-off score alone to evaluate malingering. In order to detect malingering or intentional poor performance, a multidimensional approach is necessarily required.34,35,44 As such, using the discriminant function formula introduced in South Korea by Moon and Hwang36 with the RDS cut-off score would increase the validity of the detection of malingering. However, the formula is recommended to be used with the mild TBI group because the discriminant function formula was originally designed to distinguish the mild TBI group from the malingering group. In addition, in order to detect intentional poor performance and malingering, clinicians may additionally conduct PVTs, such as the California Verbal Learning Test, the Rey-Osterrieth Complex Figure Test, the Test of Memory Malingering, and the Word Memory Test. In addition to the test data, the process that integrates all the data including the examinee's personal, medical, and psychiatric history data is essential for the detection of malingering.

The significance of this study is as follows. Few studies on PVTs in South Korea were conducted, and this study first introduced PVTs using K-WAIS-IV RDS and suggested RDS cut-off scores by group. The K-WAIS-IV is the most widely used intelligence test in South Korea and is routinely administered in psychiatric settings. RDS can be easily used to examine performance validity using WAIS without any additional tests, which is very time- and cost-effective. Moreover, calculating RDS is very simple, and even clinicians can check for poor performance in the middle of the testing. The previous studies in South Korea have not shown a fundamental epidemiological data about RDS.40 This study could be the basis for future performance validity studies.

The limitations of this study and future research directions are as follows. This study introduced RDS cut-off scores using 10% of the base rate, not using ROC analysis. The replication studies showing RDS cut-off scores through ROC analysis will increase the statistical validity. Furthermore, RDS research needs to be conducted with a variety of groups. As with previous studies, the studies to apply and examine RDS to various subject groups such as those with intellectual disabilities, patients with dementia, and chronic pain patients are needed in South Korea.18 Finally, no stand-alone-type PVTs such as the Test of Memory Malingering25 and the Word Memory Test24 are standardized in South Korea. The Korean version of standardized stand-alone-type PVTs will be able to increase the reliability and validity of detecting malingering with embedded-type PVTs together.

Acknowledgments
This study was supported by Wonkwang University in 2020.

Conflicts of Interest
The authors have no potential conflicts of interest to disclose.

Author Contributions

ORCID iDs
An Kook On http://orcid.org/0000-0002-0171-4844
Kyu-Sic Hwang http://orcid.org/0000-0002-8997-8720
Seung-Ho Jang http://orcid.org/0000-0002-3479-0552
Hye-Jin Lee http://orcid.org/0000-0002-1039-9196
Min-Jung Soh http://orcid.org/0000-0003-1033-9449
Chan-Mo Yang http://orcid.org/0000-0002-4959-7595
Sang-Yeol Lee http://orcid.org/0000-0003-1828-9992

REFERENCES
10. Meyer GJ, Archer RP. The hard science of Rorschach research: What