

K-WAIS-IV 단축형 지능검사의 IQ 추정 오차와 보정 방안*

여 환 홍¹⁾

이 윤 형[†]

단축형 지능검사에서 회귀추정 방식으로 지능을 추정할 경우 실제IQ (전체척도IQ)가 평균보다 낮은 사람들의 IQ가 과대추정 되는 경향이 있다. 이는 일정한 방향성을 지닌 체계적 오차로 단축형 지능검사의 적절성을 위협하는 요인이다. 본 연구에서는 이러한 오차를 줄이기 위해 Davis(1976)의 RTM 에 관한 수학적 모델을 응용하여 RTM 보정 공식을 고안하였다. 본 연구는 또한 이 보정 공식의 타당성을 확인하기 위해 두 개의 동질적인 낮은 지능 집단들과 이질적 정신장애 집단을 대상으로 비례배분 방식, 선형동등화 방식, 보정 전 회귀추정 방식, Ehrler 등(2019)의 보정회귀 방식과 RTM 보정 방식의 오차를 비교하였다. 그 결과 본 연구에서 제안한 보정 공식은 보정 전 회귀추정 방식보다 뚜렷하게 개선된 정확도를 보였고, 비회귀추정 방식인 비례추정 방식, 선형동등화 방식 보다 더 나은 정확도를 보였으며, Ehrler 등(2019)이 제시한 보정회귀 방식 보다도 정확도가 높은 것으로 나타났다. 아울러 보정회귀 방식을 사용한 경우에는 추정 오차의 선형성이 관찰되었지만 RTM 보정 공식을 이용하였을 때는 이러한 선형성이 사라지는 것을 확인하였다. 이러한 결과의 시사점, 의의, 한계에 대해 상세히 논의하였다.

주요어 : 지능검사, 단축형 검사, 추정IQ, 평균으로의 회귀, 오차 보정

* 이 논문은 대한민국 교육부(NRF-2020S1A3A2A02103899) 지원에 의해 수행되었음.

1) 대전충남지방병무청, 임상심리전문가, grunze83@korea.kr

† 교신저자: 이윤형, 영남대학교 문과대학 심리학과 부교수, (38541) 경상북도 경산시 대학로 280,

E-mail: yhlee01@yu.ac.kr

웍슬러 지능 검사(Wechsler-Bellevue Adult Intelligence Scale)의 출간 이후 많은 단축형 지능검사들이 개발되었다(Anastasi & Urbina, 1997). 단축형 지능검사들은 짧은 소요시간, 측정된 추정IQ와 전체척도IQ와의 높은 상관(.90 이상) 때문에 많은 분야에서 지속적으로 사용되어 왔다. 또한 단축형 지능검사의 타당성은 표준화 표본 뿐 아니라 고령(Clara & Huynh, 2003; Hoffman & Nelson, 1988), 조현병(Blyer et al, 2000; Bulzacka et al, 2016, Lim et al., 2000), 치매(Brooks & Weaver, 2005; Margolis & Taylor, 1986), 외상성 뇌손상 등 신경과 질환(Jeong et al., 2017; Ryan, Kreiner, Gontkovsky, & Umfleet, 2015), 이질적인 정신장애(So & Choi, 2019) 등 특수한 표본에 대해서도 검증되었다.

비록 몇몇 연구자들은 여전히 단축형 지능검사의 타당성에 의문을 제기하고 있고(Jung & Back, 2019; Silverstien, 1971; Kaufman & Kaufman, 2001), 단축형 지능검사 타당화 연구들(Choe et al., 2014; Hwang & Oh, 2017)에서도 진단 목적 혹은 지능을 수준별로 분류하는 목적으로 단축형 지능검사를 사용하는 것은 지양하는 것이 바람직하다고 제안하고 있지만, 불가피한 상황에서는 단축형 지능검사가 좋은 대안이 될 수 있다. 예를 들어 건강 상태 및 심리적 상태가 불안정한 급성기 환자를 대상으로 하는 경우나 장시간 인내력과 주의 집중력을 유지하기 어려운 조현병이나 뇌손상 환자 경우, 단축형 지능검사가 이들에 대한 평가 및 치료를 돕는데 유용하게 사용될 수 있다. 또한 대규모 집단을 대상으로 추적 조사하는 코호트 연구나 병역판정 검사처럼 대규모의 선별이 이루어지는 상황과 같이 현실적으로 모든 대상자에게 전체척도를 실시하는

것이 불가능한 경우에도 유용하게 사용될 수 있다.

단축형 지능검사는 전체문항 중 일부 문항을 실시하여 지능을 추정하는 방식(Resnick, 1977; Satz & Mogel, 1962)과 시간적 효율성이 높은 여러 소검사의 조합을 선택하여 사용하는 방식(Crawford et al., 2010; Doppelt, 1956; Girard et al, 2015; Silverstein, 1982; Ward, 1990)이 있는데 이 중 여러 소검사의 조합을 사용하는 것이 일반적이다.

소검사를 선별적으로 실시하는 단축형 지능검사는 소검사 점수(전체척도 기준의 연령별 환산점수)를 바탕으로 전체척도IQ를 추정하는데, 그 추정 방식은 크게 3가지로 구분된다. 우선 전체 소검사 개수를 단축형에 포함된 소검사 개수로 나눈 값에 소검사 환산점수의 합을 곱하는 비례배분 방식(proration method)이 있다. 예를 들어 Meyers 등(2013)이 제안한 WAIS-IV 단축형은 WAIS-R의 WARD 7 형에 기원한 7개 소검사를 이용하였으며 비례배분 방식을 적용하여 IQ를 도출하였다. 또한 단축형에 포함된 소검사의 합을 전체척도IQ와 동일한 단위 즉, 평균이 100이고, 표준편차가 15인 점수로 변환하는 선형동등화 방식(linear equating method)이 있는데 Crawford 등(2010)의 WISC-IV 단축형 7개 소검사(WARD 7 형)는 이 방식을 사용하였다.

단축형 지능검사의 추정 방식 중 나머지 하나는 포함된 소검사가 전체척도IQ를 추정하는 오차가 최소화 되도록 하는 회귀방정식을 사용하는 회귀추정 방식(regression formula method)이다. 앞서 설명한 비례배분, 선형동등화 방식은 전체척도IQ의 예측치라기보다 IQ 점수와 유사한 단위로 표현된 별도의 점수라고 할 수 있다(Crawford et al., 2010). 반면에

회귀추정 방식은 오차를 최소화하며 회귀계수를 통해 각 소검사의 기여도가 반영된다는 점에서 다른 추정 방식에 비해 이점이 있다 (Choe et al., 2014). 이러한 이점 때문에 회귀추정 방식을 이용한 단축형 지능검사가 널리 쓰이는데 예를 들어 Choe 등(2014)의 한국판 성인용 웨슬러 지능검사 4판(Korean Wechsler Adult Intelligence Scale-IV, K-WAIS-IV) 단축형, Hwang과 Oh(2017)의 한국판 아동용 웨슬러 지능검사 4판(Korean Wechsler Intelligence Scale for Child-IV, K-WISC-IV) 단축형은 모두 회귀추정 방식을 통해 추정IQ를 도출한다.

회귀추정 방식의 장점은 선행연구에서도 잘 나타나고 있는데 Hwang과 Oh(2017)는 표준화 집단 연구에서 4개 소검사를 바탕으로 한 회귀추정 방식이 5개 소검사 비례배분 방식, 7개 소검사 선형동등화 방식보다 더 타당하다고 제안하고 있다. 또한 이질적인 정신건강의학과 집단에 적용했을 때도 회귀추정 방식이 비례배분 방식에 비해 더 타당한 추정 IQ를 제공하는 것으로 나타났다(So & Choi, 2019).

하지만 Bulzacka 등(2016)의 조현병 환자 집단을 대상으로 한 연구, Jeong 등(2017)의 외상성 뇌손상 환자집단을 대상으로 한 연구를 살펴보면 회귀추정 방식에도 단점이 있다는 것을 잘 알 수 있다. Bulzacka 등은 조현병과 조현정동장애 환자를 대상으로 연구를 진행하였는데 대상자의 IQ 평균이 81.1이었다. 이 표본에서는 7개 소검사 단축형으로 추정IQ를 구하기 위해 비례배분 방식과 회귀추정 방식을 적용하였는데 그 결과 비례배분 방식으로 추정한 IQ는 타당한 것으로 나타났으나 회귀추정 방식은 전체척도IQ와 큰 평균 차이를 보였다. Jeong 등의 연구에서는 4개 소검사 조합

에 회귀추정 방식과 비례배분 방식을 적용했는데 그 결과 비뇌외상 환자 집단(IQ 평균 98.22)에서는 회귀추정 방식과 비례배분 방식 모두 타당도가 높았던 반면 중등도 뇌외상 환자 집단(IQ 평균 81.84)에서는 회귀추정 방식이 IQ를 유의미하게 과대추정하였다. 이러한 선행연구의 결과들은 회귀추정 방식을 임상 집단, 그 중에도 지능이 낮은 대상자들에게 적용했을 때는 정확도가 떨어진다는 것을 보여준다.

회귀추정 방식이 표본에 따라 일관되지 못한 정확성을 보이는 이유는 Jung과 Baek (2019)의 연구결과를 통해 짐작할 수 있다. 이 연구에서는 이질적인 정신장애 집단을 IQ 수준별로 나누고 표준화 집단으로부터 도출된 회귀추정 공식을 사용하여 구한 추정IQ와 전체척도IQ의 차이를 살펴보았는데 그 결과 지적장애 집단, 평균 이하 집단, 평균 이상 집단 모두에서 추정IQ와 전체척도IQ간의 차이가 유의미하였고, 평균 집단에서만 추정한 IQ가 타당한 것으로 확인되었다. 이는 회귀추정 방식을 사용하게 될 경우 지능이 높거나 낮은 대상자에서 IQ를 잘못 추정하게 될 가능성이 크다는 것을 보여준다.

이와 같은 회귀추정 방식의 문제점은 표준화 자료에서 도출된 회귀식을 낮거나 높은 수준의 지능을 지닌 집단에 적용했기 때문에 발생한 것으로 짐작된다. 이런 현상은 표준화 표본을 통해 타당도를 검증한 연구에서도 나타나는데, Choe 등(2014)은 IQ가 $\pm 1SD$ 을 넘는 경우(85 미만, 115 초과), 표준화 집단 전체에서 도출한 회귀추정식을 사용하면 IQ가 낮은 경우 과대추정 되고, 높은 경우 과소추정 된다고 밝혔다.

최근에는 표준화 집단에서 도출한 회귀추

정식을 일반적이지 않은 집단에 사용하는 것에서 발생하는 문제점을 해결하고자 하는 시도가 국내외에서 나타나고 있다. Ehrler 등(2019)은 코호트 연구에 참여하고 있는 선천성 심장질환 환자를 대상으로 새로운 보정 공식을 개발하여 적용하였다. 비슷한 시기에 국내에서도 Yeo 등(2019)은 병역판정검사에서 인지능력 저하 의심자를 대상으로 새로운 회귀추정식을 도출하였고, So와 Choi(2019)도 이질적 정신장애 집단을 대상으로 한 연구에서 각 연구표본으로부터 새로운 회귀추정식을 도출하였다.

이러한 보정은 표준화 집단으로부터 도출된 단축형에 비해 더 나은 타당도를 보인다. 하지만 이 역시도 특정 집단에서 얻어진 회귀공식을 사용하기 때문에 과적합(over fitting)의 문제를 피하기 어렵다. 즉 동질적인 집단에 사용하게 되면 정확도가 매우 높지만 IQ 분포가 다른 집단에 사용하게 되면 정확도가 떨어지는 문제가 발생할 수 있다.

표준화 집단으로부터 도출된 회귀추정 공식을 평균에서 이탈된 IQ를 지닌 집단에 적용할 경우 생기는 문제점은 선행 연구(Yeo et al., 2019; Ehrler et al., 2019)에서 평균으로의 회귀현상(Regression toward the mean; 이하 RTM)으로 설명하고 있다. RTM은 Francis Galton(1886)이 처음 제안한 것으로 상관이 완벽하지 않은 A 와 \hat{A} 가 있다면 A 라는 측정치로 \hat{A} 라는 측정치를 예측할 때, A 가 극단값을 가질 경우 \hat{A} 가 상대적으로 덜 극단적으로 예측되는 통계적 현상을 말한다. RTM의 정도는 A 와 \hat{A} 상관이 낮을수록 커지고, A 값이 평균으로부터 멀어질수록 커진다. 이런 현상은 표준화 집단에서 $\pm 1SD$ 를 벗어난 대상에 대해 회귀추정 방식으로 구한 추정IQ의 정확도가

떨어지게 되는 결과, 그리고 임상 집단에서 비례배분, 선형동등화 추정 방식보다 회귀추정 방식의 타당도가 낮은 결과를 잘 설명한다.

RTM을 다루고 있는 연구들은 주로 혈압 및 심박수(Bland, 2013; Bland & Altman, 2004), 콜레스테롤 수치(Barnett et al., 2005; Chinn & Heller, 1981; Davis, 1976)와 같이 측정시기에 따라 변동성이 큰 생물학적 측정 자료를 분석하였다. 이 연구들은 RTM 현상을 규명하기 위해 기저 측정(Y_1)과 사후 측정(Y_2)시의 변화 양상을 검토하여 차이값($y_1 - y_2$)이 기저 측정치(y_1)가 평균으로부터 멀어질수록 커진다는 것을 보여주었다(Barnett et al., 2005; Bland & Altman, 2004).

이 연구들은 또한 RTM을 정의하는 수학적 모델을 제시하고, 상황에 따라 사용할 수 있는 다양한 보정 공식을 제시하였다. 대표적으로 Davis(1976)의 수학적 모델을 공식으로 표현하면 다음과 같다.

$$RTM_{amount} = C_{(z_1)}\sigma(1 - \rho_{12}) \quad (1)$$

식 (1)에서 ρ_{12} 는 기저 측정(Y_1)과 사후 측정(Y_2) 간의 상관을 의미한다. 단축형 지능검사의 경우에는 포함된 소검사 수가 늘어날수록 전체척도와의 상관이 커진다. 따라서 RTM은 측정치 간의 상관이 낮을수록 커지며 ($1 - \rho_{12}$), 단축형 지능검사의 경우 4개 소검사보다 2개 소검사에서 RTM이 더 커지게 된다. $C_{(z_1)}$ 는 표준 정규분포 상에서 특정 지점(z_1)의 누적 분포 함수(Cumulative distribution function, $\Phi(z_1)$) 대비 확률 밀도 함수(Probability density function, $\phi(z_1)$)의 비율

(inverse Mill's ratio)이다. $C(z_1)$ 는 기저 측정 (Y_1) 특정 값(y_1)이 정규분포 상에서 관찰되지 않을 확률을 의미하며, 평균으로부터 이탈된 극단값일수록 커진다. 따라서 $C(z_1)$ 가 커질수록 즉, 극단값을 가져 관찰 확률이 희박해질수록 RTM 역시 커진다. σ 는 기저측정치 (Y_1)의 표준편차로 RTM 크기와 비례한다.

위와 같이 RTM 현상을 단축형 지능검사를 통한 지능 추정에도 동일하게 적용해 볼 수 있다. 단축형 지능검사에서 RTM 현상은 전체 척도IQ가 평균으로부터 멀어지는 정도에 비례하여 단축형 추정IQ의 오차가 커지는 형식으로 나타나게 된다. 즉, 회귀추정식이 도출된 집단의 평균을 중심으로 평균보다 클 경우 과소 추정되고, 평균보다 작을 경우 과대 추정되는 양상으로 RTM 현상이 관찰된다(Choe et al., 2014; Haynes & Atkinson, 1983; Jun, et al., 2005; Jung & Back, 2019; Yeo et al., 2020).

일반적으로 단축형 지능검사의 오차는 정보손실로 인한 불가피한 것이지만 특별한 방향성이 없는 무선적인 오차라 간주된다. 하지만 RTM 현상에서 알 수 있듯 실제 단축형 지능검사의 오차는 일정한 방향으로 과대 추정 혹은 과소 추정되는 체계적 오차(systematic error)로 인한 편향을 상당수 포함하고 있으며 이러한 체계적 오차는 수학적 보정이 필요하다. 예를 들어 추정IQ가 y_1 이고, 실제IQ가 y_2 라고 했을 때, y_1 이 평균 미만일 때 IQ가 과대추정 되고, 평균을 초과할 때 과소추정 되는 체계적 오차가 나타난다. 하지만 추정값과 실제값의 체계적 오차($y_1 - y_2$)는 추정값(y_1)이 실제값(y_2)을 예측할 때 발생할 수 있는 RTM을 제거함으로써 줄일 수 있다. 식 (1)을 이 문제에 적용해 보면 σ 는 단축형 지능검사의

표준편차이다. ρ_{12} 는 단축형 지능검사와 전체 척도 지능검사의 상관이며 z_1 은 관찰된 추정 IQ의 z점수이다. 이는 아래 식 (2)와 같이 표현될 수 있다.

$$RTM_{amount_{\hat{IQ}}} = C(z_{\hat{IQ}})\sigma_{\hat{IQ}}(1 - \rho_{FS}) \quad (2)$$

식 (2)에서 추정IQ는 \hat{IQ} 로 표기하였으며, z 점수를 구하기 위한 평균(μ) 및 표준편차(σ), 그리고 단축척도와 전체척도 지능검사의 상관 계수(ρ_{FS})는 Choe 등(2014)의 값을 사용하였다. 이는 Choe 등(2014)의 연구가 K-WAIS-IV 규준 제작을 위해 수집된 표준화 표본(N= 1,228)을 사용하여 타당화하였기 때문이다. Choe 등(2014)은 K-WAIS-IV로 개정된 이후 처음으로 단축형을 타당화하였다. 이들은 g요인과 K-WAIS-IV에서 채택된 4가지 하위요인들에 대한 요인 부하량 등을 검토하고, 시간 효율성을 고려하여 2개 소검사, 4개 소검사 단축형을 제안하였다. 이들이 제안한 단축형 검사와 전체척도와의 상관은 2개 소검사의 경우 $r = .84$, 4개 소검사의 경우 $r = .93$ 으로 상당히 높았고, 평균 차이 역시 무시할 수 있는 정도였다. 본 연구에서 사용한 RTM amount 계산 공식 및 추정IQ 보정방법은 그림1에 보다 구체적으로 제시하였다.

정보손실로 인한 오차는 크게 세 가지 방법으로 줄일 수 있다. 우선 단축형에 포함된 소검사의 수를 늘이는 것이 가장 좋은 방법일 것이다. 하지만 이러한 방법은 효율성 측면에서 문제가 있을 수 있다. 가령 10개 소검사 중 7개를 사용하는 Ward 7형의 경우 4개 소검사나 2개 소검사보다 오차가 작을 것 이지만 단축형 검사의 목적에는 부합하지 않을 수

있다. 둘째는 가장 적절한 소검사들을 단축형에 포함시키는 것이다. 오차를 최소화하는 소검사 조합을 선택하기 위해 선행연구들에서는 전체척도IQ와의 상관, 설명량(R^2), g요인 부하량 등의 방법을 사용하고 있다. 셋째는 적절한 전체척도IQ 추정 방식을 선택하는 것이다. 하지만 현재까지는 어떤 추정 방식이 오차를 최소화 할 수 있는 가장 적절한 추정 방식인지에 대해 체계적으로 살펴본 연구는 찾아보기 어렵다. 따라서 본 연구는 적절한 추정 방식을 선택하는데 초점을 맞추었으며 특히 회귀추정 방식에서 발생하는 체계적 오차를 줄이는데 초점을 두고 있다. 보다 구체적으로 본 연구에서는 Davis(1976)가 제시한 RTM의 수학적 모델을 응용하여 추정IQ의 오차를 최소화시키는 방법을 제시하고 이 방식의 정확도를 Ehrler 등(2019)의 보정 방식, 보정하지 않은 회귀추정, 비례배분, 선형동등화 방식 등과 비교해봄으로써 RTM 보정 방식의 타당성을 검증해보고자 한다.

방 법

연구 대상

본 연구는 201X년부터 201X년까지 병무청 병역판정검사 과정에서 축적된 자료를 분석하는 후향적 연구이다. 분석 자료는 한국판 웨슬러 지능검사 4판 전체척도(이하 K-WAIS-IV)이다. 본 연구에서 최초 표집한 저지능 집단은 X,XXX명으로 컴퓨터 기반 인지능력검사 1.0(한국국방연구원, 58문항)를 실시한 결과 지적능력에 문제가 있는 것으로 의심되어 선별된 이들이다. 이들은 모두 18~33세 연

표 1. 타당화 표본의 IQ 분포

지능수준	저지능 집단1	저지능 집단2	이질적 정신장애 집단
중등도 지적장애 (IQ<50)	33(5.63)	31(5.28)	-
경도 지적장애 (IQ 50-69)	377(64.33)	389(66.27)	-
경계선 지능 (IQ 80-89)	135(23.04)	124(21.12)	24(15.19)
평균 하 (IQ 80-89)	41(7.00)	43(7.33)	50(31.65)
평균 (IQ 90-110)	-	-	80(50.63)
평균 상 이상 (IQ>110)	-	-	4(2.53)

() 안 백분율

령 범위의 남성이다. 이들 중 지능지수가 평균 수준(IQ≥90)인 사람들의 자료는 분석에 포함되지 않았다. 최종적으로 선별된 저지능 집단 1,173명을 무선 할당하여 저지능 집단 1(n=586), 저지능 집단2(n=587)의 두 집단으로 나누었다. 그리고 이질적 정신장애 집단(n=158)을 별도로 표집하였다. 이 집단은 다양한 범주의 정신장애를 지닌 이질적인 표본이다. 이들은 컴퓨터 기반 인지능력검사에서 정상으로 분류된 사람들이며, 이 중에는 평균 이상의 지능을 지닌 대상자들도 상당수 포함되어 있다.

보다 구체적으로, 본 연구에서는 각 보정 방식의 정확성과 더불어 안정성을 검토하기 위해 회귀공식을 도출한 저지능 집단1과 동질적인 IQ분포를 지닌 또 다른 저지능 집단2로

나누어 타당성을 검증했다. 이는 Ehrler 등 (2019)의 타당화 방식을 참고한 것이다. 본 연구에서는 또한 이질적 정신장애 집단을 별도로 수집하여 타당화 집단에 포함시켰다. 본 연구의 보정회귀식은 저지능 집단1로부터 도출된 것인데 타당화 집단(표준화 집단 또는 임상 집단)으로부터 도출된 회귀추정 공식의 정확성이 IQ분포가 다른 집단에서도 유지되는지 분명치 않다(Yeo et al., 2020; Jung & Baek, 2019; Choe et al., 2014). 따라서 본 연구에서는 이질적인 IQ분포를 지니고 있을 가능성이 큰 임상 집단을 추가로 살펴보는 것을 통해 도출된 회귀추정 공식의 일반화 가능성을 살펴보았다.

저지능 집단1의 전체척도IQ 범위는 42~89이며, 평균 64.18, 표준편차 9.61이다. 저지능 집단2의 전체척도IQ 범위는 40~89이며, 평균 63.43, 표준편차 9.76이다. 그리고 이질적 정신장애 집단은 정신병적 장애, 양극성 장애, 주요우울 장애 및 그 밖에 기분 장애, 신경증적 장애 등을 겪고 있는 자들 중 경계선 지능 및 지적 장애가 아닌 자들이다. 이들의 전체척도 IQ 범위는 70~120이며, 평균 90.26, 표준편차 10.43이다.

분석 방법

한국어판 WAIS-IV 단축형 타당화 연구(Choe et al., 2014)에서는 4개 소검사 단축형(information; 이하 IN, arithmetics; 이하 AR, matrix reasoning; 이하 MR, coding; 이하 CD)과 2개 소검사 단축형(IN, AR)을 제시하고 있다. 이에 따라 수집된 자료의 전체척도에서 IN, AR, MR, CD 4개 소검사를 추출하여 Choe 등 (2014)에서 제시한 회귀방정식을 통해 IQ를

계산하였다.

본 연구에서는 이 회귀추정식으로 구한 추정IQ는 RTM로 인한 체계적 오차로 인해 낮은 지능 집단의 실제IQ보다 과대추정된 수치일 것이라 예상하였다. 따라서 본 연구에서는 Davis(1976)의 RTM 수학적 모델을 참고하여 단축형 지능검사에 적용할 수 있는 RTM amount 공식을 고안하여 체계적인 오차를 통제함으로써 추정IQ를 보정하고자 하였다.

이 보정 공식은 Choe 등(2014)이 제시하고 있는 전체척도와 단축척도의 평균, 표준편차, 그리고 IQ 85 미만(저지능 집단)에서 전체척도와 단축형 간의 상관계수를 활용한다. 공식에 대한 보다 자세한 내용은 그림 1에 제시하였다.

보정 공식의 타당성을 살펴보기 위해 RTM이 보정된 추정IQ, 기존의 Choe 등(2014)의 추정IQ, 그리고 Ehrler 등(2019)의 보정회귀 방법으로 보정된 추정IQ, 그리고 회귀추정 방식이 아닌 비례배분 방식, 선형동등화 방식으로 계산된 추정IQ를 계산하고 각 추정IQ들과 전체척도IQ와의 차이를 분석 및 비교하였다. 추정IQ와 전체척도IQ와의 차이는 평균 제곱근 오차(Root Mean Square Error; RMSE)와 평균 절대 오차(Mean Absolute Error)로 살펴보았으며 이들을 구하는 공식은 식 (3)과 같다. 식에서 단축형으로 추정된 IQ를 SFIQ, 전체척도 IQ를 FSIQ로 표기하였다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (SFIQ_i - FSIQ_i)^2} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |SFIQ_i - FSIQ_i|$$

RTM amount^a $= (1 - \rho_{FS}) \times \sigma_{\hat{I}Q} \times c(z_{\hat{I}Q})$ $= (1 - \rho_{FS}) \times \sigma_{\hat{I}Q} \times \frac{PDF(z_{\hat{I}Q})}{CDF(z_{\hat{I}Q})} \quad \hat{I}Q < \mu_{\hat{I}Q}$ $= (1 - \rho_{FS}) \times \sigma_{\hat{I}Q} \times \frac{PDF(z_{\hat{I}Q})}{(1 - CDF(z_{\hat{I}Q}))} \quad \hat{I}Q > \mu_{\hat{I}Q}$	$\rho_{FS_{2\text{subtests}}}$ $= .84^b, .53^b(\text{if } IQ < 85), .54^b(\text{if } IQ > 115)$ $\rho_{FS_{4\text{subtests}}}$ $= .93^b, .75^b(\text{if } IQ < 85), .72^b(\text{if } IQ > 115)$ $\mu_{\hat{I}Q_{2\text{subtest}}} = 100.24^b$ $\mu_{\hat{I}Q_{4\text{subtest}}} = 100.23^b$ $\sigma_{\hat{I}Q_{2\text{subtest}}} = 12.54^b$ $\sigma_{\hat{I}Q_{4\text{subtest}}} = 13.85^b$
$\text{RTM corrected } \hat{I}Q$ $= \hat{I}Q - \text{RTM amount} \quad \hat{I}Q < \mu_{\hat{I}Q}$ $= \hat{I}Q + \text{RTM amount} \quad \hat{I}Q > \mu_{\hat{I}Q}$	$PDF(z_{\hat{I}Q}) : \text{추정IQ의 표준정규분포 상의 확률 밀도함수}$ $CDF(z_{\hat{I}Q}) : \text{추정IQ의 표준정규분포 상의 누적 분포함수}$
	$*\hat{I}Q = \text{단축형으로 추정된 IQ}$ $*z_{\hat{I}Q} = \text{추정IQ의 } z \text{ 점수}$ $*\rho_{FS_{2\text{subtests}}} = \text{2개 소검사 단축형과 전체척도의 상관}$ $*\rho_{FS_{4\text{subtests}}} = \text{4개 소검사 단축형과 전체척도의 상관}$ $*\mu_{\hat{I}Q_{2\text{subtest}}} = \text{2개 소검사의 추정IQ 평균}$ $*\mu_{\hat{I}Q_{4\text{subtest}}} = \text{4개 소검사의 추정IQ 평균}$ $*\sigma_{\hat{I}Q_{2\text{subtest}}} = \text{2개 소검사의 추정IQ의 표준편차}$ $*\sigma_{\hat{I}Q_{4\text{subtest}}} = \text{4개 소검사의 추정IQ의 표준편차}$

- a. 단축형 지능검사 RTM 보정 계산기(<https://trello.com/c/DoWEwrKI>) 및 부록 참고
 b. 표준화 집단(N=1,228)의 전체척도IQ와 추정IQ의 평균, 표준편차, 전체척도 및 단축척도의 상관계수. Choe 등(2014)에서 추출

그림 1. 단축형 지능검사 추정IQ의 RTM amount 계산공식과 보정방법

아울러 본 연구에서는 RTM이 보정된 추정 IQ, Choe 등(2014)의 추정IQ, Ehrler 등(2019)의 방식으로 보정된 추정IQ, 비례배분 방식으로 계산된 추정IQ, 선형동등화 방식으로 계산된 추정IQ 별로 오차 크기를 비교할 수 있도록 RMSE를 각 집단의 전체척도 IQ의 평균으로 나누는 정규화 평균 제곱근 오차(Normalized Root Mean Square Error; NRMSE), 실제값 대비

절대오차 비율의 평균인 평균 절대 오차 백분율(Mean Absolute Percentage Error; MAPE)을 계산하였다. 또한 각 추정 방법별로 체계적 오차가 나타나는지 여부를 살펴보기 위해 각 추정IQ와 전체척도IQ와의 오차를 산포도를 통해 검토하고, 전체척도IQ의 오차변량에 대한 R²을 살펴보았다.

표 2. IQ 추정에 사용된 각 추정 및 보정공식

추정방식	공식	참고문헌
보정 전 ^c (회귀추정)	four-SFIQ ^a = 39.021+1.569*IN+1.477*MR+1.536*AR+1.4905*CD two-SFIQ ^b = 54.762+2.330*IN+2.151*AR	Choe 등(2014)
보정회귀 ^d	four-SFIQ _{Ehrler} = four-SFIQ*0.92-1.257 two-SFIQ _{Ehrler} = two-SFIQ*0.952-10.947	Ehrler 등(2019)
RTM 보정	four-SFIQ _{RTMcorrected} = four-SFIQ±RTM amount two-SFIQ _{RTMcorrected} = two-SFIQ±RTM amount	Davis(1976)
비례배분	four-SFIQ _{proportion} = (IN+AR+MR+CD)*10/4 two-SFIQ _{proportion} = (IN+AR)*10/2	Meyers 등(2013)
선형동등	four-SFIQ _{LE} = SD _{new} /SD _{old} *(IN+AR+MR+CD-40)+100 two-SFIQ _{LE} = SD _{new} /SD _{old} *(IN+AR-20)+100	Crawford 등(2010)

- a. 4개 소검사 = 상식(IN), 산수(AR), 행렬추론(MR), 기호쓰기(CD),
- b. 2개 소검사 = 상식(IN), 산수(AR)
- c. Choe 등(2014)에서 도출한 공식을 그대로 사용
- d. 추정IQ가 전체척도IQ를 예측하는 단순회귀분석을 통해 얻어졌으며 본 연구 표본 저지능 집단1에서 도출

결 과

IQ 추정 공식 및 보정 공식의 정확성 검토

2개 소검사, 4개 소검사 단축형의 추정IQ를 보정하기 위해 개발된 보정 공식의 타당성과 안정성을 알아보려고 저지능 집단1, 저지능 집단2 그리고 이질적 정신장애 집단의 세 검증 집단을 대상으로 기존 회귀추정식(Choe et al., 2014), Ehrler 등(2019)의 보정회귀, 본 연구에서 제안한 RTM 보정 공식을 활용한 추정IQ의 정확성을 비교하였다. 아울러 비회귀추정 방식인 비례배분 방식, 선형동등화 방식을 이용한 추정IQ와도 비교하였다. 또한 타당도 집단 별로 각 추정IQ 정확성이 달라지는지도 살펴보았다. 이를 위해 5가지 추정 방식 및

보정 방식의 추정IQ를 계산하고, 평균 및 전체척도IQ와의 상관을 살펴보았으며, 오차(추정IQ - 전체척도IQ)크기를 비교하였다.

낮은 지능 집단의 각 추정IQ와 전체척도IQ 간의 상관계수는 2개 소검사에서는 $r = .73 \sim .75$ 이며, 4개 소검사에서는 $r = .87 \sim .88$ 이다. 상관은 단축형에 포함된 소검사 개수에 따라 차이를 보였지만 추정 방식에 따라서는 비슷한 수준으로 나타났다. 예외적으로 이질적 정신장애 집단에서는 RTM 보정방식으로 보정된 경우 2개 소검사 $r = .83$, 4개 소검사 $r = .92$ 로 다른 추정 방식에 비해 상대적으로 높은 상관을 보였다.

각 추정 방식의 타당성은 전체척도IQ와의 오차를 얼마나 최소화 하는가로 규정될 수 있다. 따라서 정확성이 높은 추정 방식 및 보정

표 3. 각 보정 공식으로 추정된 IQ(SFIQ)와 전체척도IQ(FSIQ) 간의 상관, 평균 차이 및 오차 비교

추정 방식	r_{FS}	SFIQ M(SD)	Mean Diff	Cohen's <i>d</i>	RMSE (NRMSE)	MAE (MAPE)	Mean of error (<i>n</i>)	
							error \geq 0	error < 0
2개 소검사								
낮은 지능 집단 1 (n=586) FSIQ M(SD) = 64.18 (9.61)								
보정 전	.75	78.53(7.12)	14.35	2.26	15.69(.24)	14.42(24%)	14.58(578)	-2.57(8)
보정회귀	.75	63.81(6.82)	-0.37	.06	6.38(.10)	5.09(8%)	4.79(289)	-5.38(297)
RTM 보정	.77	66.02(10.24)	1.84	.27	7.06(.11)	5.60(9%)	6.06(360)	-4.89(226)
비례배분	.75	53.37(15.70)	-10.81	1.02	15.16(.24)	12.97(21%)	6.40(99)	-14.31(487)
선형동등	.75	73.95(8.77)	9.77	1.49	11.77(.18)	10.17(17%)	10.66(548)	-3.11(38)
낮은 지능 집단 2 (n=587) FSIQ M(SD) = 63.43 (9.76)								
보정 전	.74	77.83(7.02)	14.40	2.19	15.85(.25)	14.52(25%)	14.79(574)	-2.73(13)
보정회귀	.74	63.15(7.08)	-0.28	.04	6.63(.10)	5.29(8%)	4.87(302)	-5.74(285)
RTM 보정	.75	65.03(10.16)	1.60	.23	7.25(.11)	5.74(9%)	5.96(361)	-5.37(226)
비례배분	.73	51.83(15.56)	-11.60	1.08	15.78(.25)	13.27(22%)	5.46(90)	-14.68(497)
선형동등	.73	73.09(8.70)	9.66	1.41	11.82(.19)	10.31(17%)	10.86(540)	-4.07(47)
이질적 정신장애 집단 (n=158) FSIQ M(SD) = 90.26 (8.99)								
보정 전	.76	95.10(9.18)	-4.84	.71	8.38(.09)	6.88(8%)	7.47(124)	-4.75(34)
보정회귀	.76	79.59(8.74)	10.67	1.55	12.64(.14)	11.04(12%)	3.69(8)	-11.44(150)
RTM 보정	.83	91.79(12.77)	-1.53	.22	7.25(.08)	5.62(6%)	6.28(90)	-4.76(68)
비례배분	.76	90.06(20.29)	.20	.01	14.05(.16)	11.12(12%)	11.82(73)	-10.52(85)
선형동등	.76	94.45(11.34)	-4.19	.55	8.64(.10)	6.88(8%)	7.41(118)	-5.31(40)
4개 소검사								
낮은 지능 집단 1 (n=586)								
보정 전	.87	70.87(8.70)	6.68	1.41	8.20(.13)	7.13(12%)	7.51(539)	-2.76(47)
보정회귀	.87	63.94(8.00)	-0.24	.05	4.76(.07)	3.80(6%)	3.46(301)	-4.15(285)
RTM 보정	.88	62.35(10.84)	-1.83	.36	5.54(.09)	4.43(7%)	3.46(220)	-5.01(366)
비례배분	.87	52.67(14.45)	-11.51	1.49	13.87(.22)	12.04(20%)	3.00(51)	-12.90(535)
선형동등	.87	68.79(9.53)	4.61	.95	6.72(.10)	5.57(9%)	6.07(491)	-2.95(95)
낮은 지능 집단 2 (n=587)								
보정 전	.87	70.05(8.83)	6.62	1.37	8.19(.13)	7.10(12%)	7.49(538)	-2.65(49)
보정회귀	.87	63.18(8.12)	-0.24	.05	4.78(.08)	3.75(6%)	3.44(302)	-4.10(285)
RTM 보정	.88	61.34(11.00)	-2.09	.40	5.68(.09)	4.52(7%)	3.76(192)	-4.88(395)
비례배분	.87	51.31(14.67)	-12.12	1.55	14.38(.23)	12.64(21%)	4.18(39)	-13.23(548)
선형동등	.87	67.90(9.67)	4.47	.90	6.66(.11)	5.44(9%)	5.95(490)	-2.85(97)
이질적 정신장애 집단 (n=158)								
보정 전	.87	90.19(9.54)	.07	.01	5.10(.06)	3.99(4%)	3.60(86)	-4.46(72)
보정회귀	.87	81.72(8.78)	8.54	1.68	9.94(.11)	8.66(9%)	1.52(6)	-8.94(152)
RTM 보정	.92	88.57(9.64)	1.69	.35	5.58(.06)	4.02(4%)	2.66(69)	-5.07(89)
비례배분	.87	84.57(15.75)	5.69	.64	10.10(.11)	8.10(9%)	4.55(42)	-9.39(116)
선형동등	.87	89.83(10.39)	.43	.08	5.25(.06)	4.04(4%)	3.56(80)	-4.52(78)

방식을 알고자 RMSE, MAE로 오차의 크기를 구하였다. 그리고 이들의 규모 의존적(scale dependent) 특성을 보완하기 위해 RMSE를 정규화한 NRMSE, 절대오차를 전체척도IQ 대비 비율로 나타낸 MAPE를 사용하여 상호 비교하였다. Lewis(1982)는 MAPE의 해석기준을 제시하고 있는데, 10% 미만은 ‘매우 정확한 예측’, 20%는 ‘좋은 예측’, 20~50%는 ‘합리적 수준의 예측’, 50%를 초과하면 ‘부정확한 예측’이다. 본 연구에서는 ‘매우 정확한 예측’을 의미하는 MAPE 10% 미만으로 기준을 설정하여 각 단축형 지능검사 및 추정 방식의 타당성을 검토하였다. 그 결과는 표 3과 같다.

보정 전 회귀추정방식의 경우 저지능 집단 1, 2에서 2개 소검사, 4개 소검사 모두 MAPE가 10%를 초과했고, 특히 2개 소검사의 경우 20%를 초과했다. 반면 이질적 정신장애 집단에서는 MAPE가 2개 소검사에서 8%, 4개 소검사에서 4%로 높은 정확도를 보였다. 보정 회귀공식을 사용한 경우 2개, 4개 소검사 모두, 낮은 지능 집단 모두에서 MAPE가 10% 미만으로 높은 정확도를 보였던 반면 이질적 정신장애 집단에서는 2, 4개 소검사 모두 보정 전보다 RMSE(MAE)가 각각 12.64(11.04), 9.94(8.66)로 두 배 가량 오차가 커졌다. 그에 반해 RTM 보정방식은 2개, 4개 소검사 및 세 타당화 집단 모두에서 MAPE가 10% 미만이었으며, 각 타당도 집단에서 오차가 2개 소검사 RMSE 7.06~7.25, MAE 5.60~5.74, 4개 소검사 RMSE 5.54~5.68, MAE 4.02~4.52으로 타당화 집단 모두에서 오차 크기가 비슷하였다.

회귀추정 방식이 아닌 비례배분 방식과 선형동등화 방식의 오차를 살펴보면 비례배분 방식의 경우 4개 소검사, 이질적 정신장애 집단에 한정해서만 MAPE가 9%로 높은 정확도

를 보였다. 선형동등화 방식은 단축형이 4개 소검사를 포함할 경우에는 모든 집단에서 높은 정확도를 보였던 반면 2개 소검사의 경우에는 이질적 정신장애 집단에서만 MAPE 8%로 높은 타당도를 보였다. 또한 동일조건에서 RTM 보정 방식은 MAPE 6%으로 더 높은 정확도를 나타냈다.

전체척도IQ와 각 추정 방식으로 구한 단축형 추정IQ의 평균 차이에 대해 유의도 검증(대응 T검증)을 실시한 결과 보정회귀 방식의 경우 2개, 4개 소검사, 두 낮은 지능 집단에서 p 값이 .05를 초과하여 전체척도IQ와 평균 차이가 유의미하지 않았다(2개 소검사 & 낮은 지능 집단1 $t(585) = -1.39, p = .16$, 2개 소검사 & 낮은 지능 집단2 $t(586) = -1.02, p = .31$, 4개 소검사 & 낮은 지능 집단1 $t(585) = -1.23, p = .22$, 4개 소검사 & 낮은 지능 집단2 $t(586) = -1.11, p = .27$). 보정 전 회귀추정 방식의 경우 4개 소검사로 이질적 정신장애 집단에 적용했을 때 p 값이 .05를 초과하였으며($t(157) = .17, p = .86$), 비례배분 방식은 2개 소검사, 이질적 정신장애 집단에서만 p 값이 .05 초과하였고($t(157) = .18, p = .86$), 선형동등화 방식은 4개 소검사로 이질적 정신장애 집단에 적용했을 때 p 값이 .05를 초과하였다($t(157) = 1.03, p = .30$). RTM 보정방식의 경우 모든 조건에서 오차가 작았지만, 전체척도IQ와 평균 차이는 모든 조건에서 유의미하였다. 그리고 이질적 정신장애 집단에서 보면 2개 소검사 및 비례배분 방식이 전체척도IQ와 매우 작은 평균 차이($Mean\ diff = .20, d = .01$)를 보였지만 다른 추정 방식에 비해 오차는 RMSE = 14.05로 가장 큰 것으로 나타났다. 게다가 위와 동일한 조건에서 2개 소검사보다 오히려 4개 소검사서 평균 차이가

켰다($Mean\ diff = 5.69, d = .64$).

각 IQ 추정 방식의 오차 편향

각 추정 방식이 IQ를 과대 혹은 과소추정 되는 정도를 상세히 살펴보기 위해 오차가 0 이거나 양수값을 가진 사례($error \geq 0$)와 음수 값을 가진 사례($error < 0$)를 나누어 오차의 평균, 그리고 각 조건에 포함된 사례의 비중을 살펴보았다. 표 3의 보정 전 회귀추정 방식을 살펴보면 저지능 집단에서는 두 가지 단축형 모두에서 양수 오차의 평균이 현저히 높았고 포함된 사례수도 이 조건에 편중되어 있어 과대추정 되는 현상이 뚜렷하게 드러나고 있다 (2개 소검사 $ME(n/N) = 14.58(578/586)$, 14.79 (574/587), 4개 소검사 $ME(n/N) = 7.51(539/586)$, 7.49(538/587)). 반면 이질적 정신장애 집단에서는 이러한 편향이 2개 소검사에서만 나타났다($ME(n/N) = 7.47(124/158)$). 보정회귀 방식의 경우 낮은 지능 집단에서는 오차의 크기와 사례가 편향되는 현상이 보정되었으나 이질적 정신장애 집단에서는 오히려 음수 오차의 크기가 더욱 커지고, 이 조건에 사례수

도 편중되어 과소추정 되는 편향이 발생하였다(2개 소검사 $ME(n/N) = -11.44(150/158)$, 4개 소검사 $ME(n/N) = -8.94(152/158)$). 반면에 RTM 보정 방식의 경우 보정회귀 방식을 낮은 지능 집단에 적용한 경우에 비하면 양수 오차가 그 크기와 사례수가 더 많긴 하지만 과대추정 편향은 보정 전에 비해 상당히 줄어들었으며(2개 소검사 $ME(n/N) = 5.96(361/587) \sim 6.06(360/586)$, 4개 소검사 $ME(n/N) = 3.46(220/586) \sim 3.76(192/587)$), 이질적 정신장애 집단에서도 비슷한 양상이 관찰되었다(2개 소검사 $ME(n/N) = 6.28(90/158)$, 4개 소검사 $ME(n/N) = 2.66(69/158)$). 비례배분 방식의 경우에는 추정방식 및 단축형 종류에 관계없이 과소 추정되는 방향으로 편향되어 있었고, 선형동등화 방식의 경우에는 4개 소검사 및 이질적 정신장애 집단을 제외하고는 모든 조건에서 과대추정 편향을 보였다.

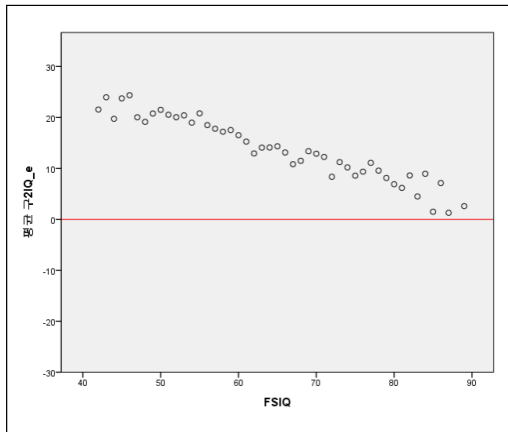
두 보정 방식의 RTM 통제

추가적으로 각 보정 방식이 RTM으로 인한 체계적 오차를 통제했는지 알아보기 위해 전

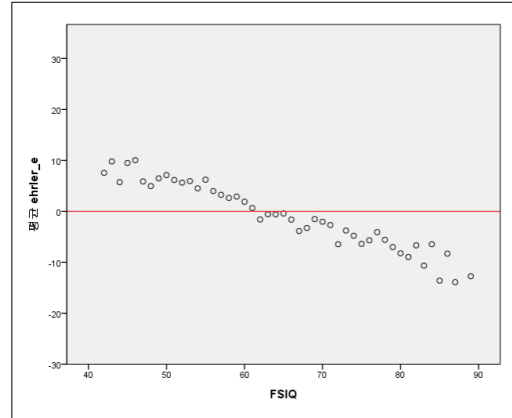
표 4. 단축형 지능검사의 추정IQ가 가지는 RTM 오차와 두 보정 방법 적용 후 오차의 체계성 변화

단축형	RTM amount M(SD)			보정 전후 R ^{2a}		
	저지능 집단1	저지능 집단2	이질적 정신장애	저지능1	저지능2	이질적 정신장애
2개 소검사	12.51(3.21)	12.79(3.22)	6.19(2.64)	.45 → .07 ^b → .51 ^c	.48 → .09 ^b → .53 ^c	.26 → .00 ^b → .32 ^c
4개 소검사	8.52(2.21)	8.68(2.26)	2.65(2.50)	.19 → .00 ^b → .31 ^c	.18 → .00 ^b → .31 ^c	.17 → .07 ^b → .29 ^c

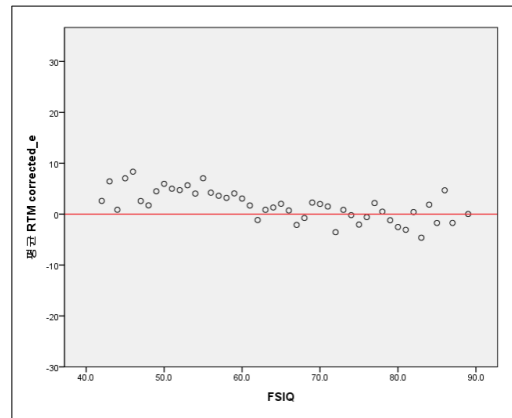
- a. 추정IQ 오차 변량에 대한 전체척도IQ의 설명량
- b. RTM 보정방법 적용
- c. Ehrler 등(2019)의 보정 회귀식 적용



2.1 전체척도IQ에 따른 추정IQ의 오차



2.2 보정회귀 공식으로 보정 후 전체척도IQ에 따른 추정IQ의 오차



2.3 RTM 보정 후 전체척도IQ에 따른 추정IQ의 오차

그림 2. 전체척도IQ에 따른 각 추정IQ의 오차 요약 산포도

체척도IQ에 따라 추정오차(추정IQ-전체척도IQ)가 선형성을 보이는지 알아보았다. 그 결과는 표 4, 그림 2에서 확인할 수 있다. 각 소검사 및 타당화 집단별 RTM amount를 보면 4개 소검사 보다는 2개 소검사, 이질적 정신장애 집단보다 저지능 집단이 더 큰 것을 확인할 수 있다.

전체척도IQ에 따라 추정오차가 변화하는지

를 살펴보기 위해 추정오차를 전체척도 IQ가 설명하는 설명량(R^2)을 살펴보면 저지능 집단1에서 보정 전 2개 소검사 추정IQ의 경우 오차 변량의 45%를 전체척도IQ가 설명하였다. 보정회귀 공식으로 보정한 후에도 전체척도 IQ가 오차 변량의 51%를 설명했다. 반면 RTM 보정 방법을 적용한 경우에는 전체척도 IQ의 오차 변량에 대한 R^2 은 7%에 불과하였

다. 즉, RTM 보정 방식의 오차변량에 대한 설명량은 현격히 줄어든 반면 보정회귀 공식의 경우 오히려 보정 후의 R^2 이 약간 증가했다는 것을 확인할 수 있다. 이러한 결과는 저지능 집단², 이질적 정신장애 집단에서도 마찬가지로 보정회귀공식을 적용하였을 때는 체계적인 오차가 통제되지 않았다.

그림 2에 제시된 전체척도IQ에 따른 단축형IQ 오차의 요약 산포도를 살펴보면 이러한 사실을 보다 쉽게 알 수 있다. 그림 2에서 보정회귀 공식의 요약 산포도를 보면 전체척도IQ가 낮아짐에 따라 오차가 증가하고 있으며, 전체척도IQ의 평균 인근, 오차가 0인 지점을 표시한 수평선을 교차하여 오차의 부호가 양에서 음으로 바뀌는 것을 확인할 수 있다. 즉 평균보다 작아지는 정도에 비례하여 과대 추정되는 정도가 커지고, 평균보다 커지는 정도에 비례하여 과소 추정되는 정도가 커지는 선형성을 지니고 있는 것을 확인할 수 있다. 그에 반해 RTM 보정 방법의 오차는 기울기가 수평에 가까우며, 오차가 0인 수평선 주변에서 무선적으로 퍼져 있는 것을 확인할 수 있다.

논 의

본 연구는 기존 단축형 지능검사(Choe et al, 2014)를 평균보다 낮은 지능을 가진 대상자(IQ 85 미만)에게 사용하면 추정IQ가 과대추정되는 체계적 오차를 보정하는 방법을 고안하고자 수행되었다. 단축형 지능검사는 건강한 사람들이나 우수한 사람을 선별하는 목적으로 실시되기 보다는 지적능력의 문제가 의심되는 임상 집단에 더욱 빈번히 활용된다.

따라서 지능이 낮은 사람들에게 단축형 지능검사를 시행했을 때 나타나는 체계적 오차를 제거하는 것은 매우 중요하다고 할 수 있다. 이런 문제를 해결하고자 하는 시도는 Ehrler 등(2019)의 연구에서 이루어졌다. 이들이 제시한 보정 방법은 단축형 지능검사를 적용할 집단(선천성 심장병)의 자료를 사용하여 단축형의 추정IQ가 전체척도IQ를 예측하는 별도의 회귀식을 만들어 활용하는 것이다. 이러한 방법은 적용할 집단에 대해서는 오차를 최소화시키는 방법이지만 RTM으로 인한 체계적 오차를 보정할 수 있는 방법은 아니다.

따라서 본 연구에서는 Davis(1976)의 RTM 수학적 모델을 참고하여 기존 단축형 지능검사의 추정IQ를 보정하는 방법을 고안하였다. 본 연구에서는 이 RTM 보정방법의 타당성을 확인하기 위하여 낮은 지능 집단을 동질적인 IQ 분포를 지닌 두 타당화 집단으로 구분하고 추가로 이질적 정신장애 집단을 포함하여 총 세 타당화 집단을 구성하였다. 이 후 이렇게 구성된 세 집단을 대상으로 보정 전 회귀 추정방식, Ehrler 등(2019)의 보정회귀공식, 비회귀추정방식인 비례배분, 선형동등화 방식과 본 연구에서 제안하는 RTM 보정 방식과의 타당성을 비교하였다. 아울러 RTM으로 인한 체계적 오차가 통제되었는지 확인하기 위하여 전체척도IQ와 추정IQ의 오차 간에 선형성을 살펴보았다. 본 연구에서 확인된 주요 결과들을 요약하면 아래와 같다.

첫째, 보정 전 회귀추정 방식은 낮은 지능 집단의 경우 2개 소검사 및 4개 소검사 모두에서 MAPE 20%를 초과하여 정확한 예측을 하지 못했다. 반면 이질적 정신장애 집단의 경우에는 2개 소검사 및 4개 소검사 모두에서 MAPE 10% 미만으로 높은 정확도를 보였다.

둘째, 보정회귀 방식은 낮은 지능 집단에 MAPE 10% 미만으로 높은 정확도를 보였지만 이질적 정신장애 집단에 적용할 경우 오히려 두 종류의 단축형 모두에서 오차 크기(RMSE, MAE)가 두 배 가량 증가하였다. 또한 오차는 일정한 편향을 보이는데, 이질적 정신장애 집단 전체 158명 중 150명에서 전체지능보다 평균 -11.44점 과소추정하였다. 셋째, RTM 보정 방식은 모든 집단과 조건에서 MAPE 10% 미만으로 높은 정확도를 보였다. 타당화 집단 별로 오차 크기를 비교해 보면 2개 소검사에서 저지능 집단1, 저지능 집단2, 이질적 정신장애 집단 순으로 RMSE가 7.06, 7.25, 7.25 였으며, 4개 소검사에서는 5.54, 5.68, 5.58로 각 조건에서 비슷한 수준의 오차를 보였다. 넷째, 비례배분 방식의 경우 이질적 정신장애 집단 및 4개 소검사에서만 MAPE가 10% 미만이었으며, 그 밖의 조건에서는 모두 이 기준을 초과하여 추정방식 중 가장 낮은 타당도를 보였다. 다섯째, 선형동등화 방식의 경우 4개 소검사에서 모든 타당화 집단에서 MAPE가 10% 미만의 높은 정확도를 보였지만, 2개 소검사에서는 이질적 정신장애 집단에서만 MAPE가 10% 미만으로 높은 정확도를 보였다.

또한 단축형 추정IQ의 오차가 전체척도IQ와 선형적 관계를 가지는 현상, 바꿔 말하면 전체척도IQ가 평균으로부터 멀어질수록 IQ를 과대 추정하는 정도가 체계적으로 커지는 오차가 관찰되는지 여부를 확인하고, 각 보정식이 이를 통제하는지를 확인하였다. 그 결과 보정 전 회귀추정 방식과 보정회귀 방식에서 전체척도IQ가 평균으로부터 멀어지는 정도에 비례해서 추정IQ의 오차가 커지는 선형성을 확인했으며, 이러한 선형성이 4개 소검사보다 2개 소검사에서 더욱 강해지는 것을 확인하였

다. 보정회귀 방식의 경우에도 역시 오차와 전체척도IQ 간의 선형성이 보정 전후가 비슷하여 체계적 오차가 통제되었다고 보기는 어렵다. 반면에 본 연구에서 제안한 RTM 보정 방식을 적용한 경우에는 모든 조건에서 오차의 선형성이 나타나지 않아 RTM으로 인한 체계적 오차를 성공적으로 통제할 수 있었다.

선행연구에서는 Resnick과 Entin(1971)가 제안한 방법을 준용하여 추정IQ와 전체척도IQ의 평균 차이 검증을 실시해 유의도, 효과크기를 확인하는 타당도를 검증하였다. 하지만 전체척도IQ와의 평균 차이가 작다고 해서 추정IQ의 정확도가 높은 것은 아니기 때문에 타당도를 검증하기 위해 이 방법을 사용하는 것은 문제가 있다. 본 연구 결과에서 RTM 보정 방식이 단축형의 종류 및 타당화 집단을 통틀어 높은 정확성을 보였음에도 전체척도와의 평균 차이에 대한 효과크기는 다른 추정방식보다 큰 것으로 나타났다. 또한 비례배분 방식의 경우 2개 소검사 및 이질적 정신장애 조건에서 추정IQ와 전체척도IQ 간에 평균 차이의 유의도 검증 결과 p 값이 .05를 초과하는 것으로 나타났으나 동일조건 다른 추정 방식에 비해 오차는 가장 큰 것으로 관찰된다. 무엇보다 이질적 정신장애 집단에 비례배분 방식을 적용하였을 때 2개 소검사가 정보손실로 인한 오차가 클 수밖에 없음에도 4개 소검사로 추정한 IQ와 전체척도IQ와의 평균 차이 검증 결과 효과크기가 .64로 ‘매우 큰 수준’으로 나타나 비례배분 방식을 적용할 경우 4개 소검사가 오히려 2개 소검사보다 타당도가 낮다는 모순되고 혼란스러운 결과가 나타났다. 이는 기존 타당도 검증 방식의 문제점을 드러내는 결과로 표준화 자료로부터 최소자승법으로 도출한 회귀추정 방식이 비례추정 방식 및

선형동등화 방식과 비교해 타당도가 높다(e.g., Choe et al., 2014; Hwang & Oh, 2017; Jun et al., 2003) 혹은 타당도가 낮다(e.g., Bulzacka et al., 2016; Meyers et al., 2013; So & Choi, 2019)는 상반된 선행연구 결과가 나타난 배경 중 하나로 꼽을 수 있다. 이러한 이유로 본 연구에서는 타당도를 검증하기 위해 예측의 오차(혹은 정확성)를 나타내는 RMSE, MAE를 사용하였다.

선행연구들에서 상반된 주장들이 상존하는 것은 위와 같은 방법론적 문제점 말고도 다른 이유가 존재한다. 이는 회귀추정 방식이 가진 본질적인 맹점인 RTM 현상이다. 이 현상은 표준화 집단에서 도출된 단축형 회귀추정 공식을 임상집단(지적장애 혹은 신경(발달)인지장애)에 적용할 경우 타당도가 낮고, IQ가 과대추정 된다는 결과들이 지속적으로 보고된 원인으로 보인다. 이러한 문제점을 해결하기 위해 So와 Choi(2019), 그리고 Yeo 등(2019)은 그들이 단축형 지능검사를 적용하고자 하는 집단인 이질적인 정신장애 집단, 병역판정검사에서 낮은 지능이 의심되는 집단에서 회귀추정식을 다시 도출한 방법을 사용하였다. 이 방식을 본 연구에서 보정회귀 방식으로 명명하였는데, 이런 방법 역시 회귀추정 방식이 지닌 본질적인 한계를 벗어나기 어렵다. 이 방법은 회귀추정식을 도출한 집단 혹은 비슷한 IQ 분포를 지닌 집단에 적용되었을 때는 높은 타당도를 보이지만 IQ 분포가 다른 집단에 적용되면 타당도가 낮아지는 과적합의 문제가 있으며, 동일집단이라 하더라도 그 집단 내 양극단 값에서 오차가 커지는 문제가 여전히 존재한다. 본 연구의 결과에서도 보정회귀 방식은 그 보정회귀식을 도출한 낮은 지능 집단에서만 타당도가 높고, IQ의 분포가

다른 이질적 정신장애 집단에서는 타당도가 낮았으며, 추정IQ의 오차는 여전히 전체적도 IQ와 뚜렷한 선형성을 지녔다.

이러한 RTM 현상으로 나타나는 체계적 오차를 극복하고자 본 연구에서는 Davis(1976)가 제안한 RTM 수학적 모델을 응용하여 RTM 보정 공식을 고안하였다. 이 보정 공식을 적용하여 단축형 추정IQ를 보정한 결과 모든 타당화 집단에서 비슷한 정도로 타당도가 높았으며 추정IQ의 오차는 전체적도IQ와 선형성을 가지지 않는 것으로 나타났다. 이런 결과는 본 연구에서 제안하는 RTM 보정방식이 체계적 오차 및 특정 집단에서만 타당도가 높은 과적합 문제를 해결했다는 것을 의미한다.

본 연구 결과를 종합하면 다음과 같은 결론을 내릴 수 있다. 지능이 낮을 것으로 예상되는 대상자에게 단축형 지능검사를 실시할 때 본 연구에서 제안한 보정 공식을 적용함으로써 추정IQ의 정확성을 증가시킬 수 있다. 특히 2개 소검사와 같은 짧은 소검사로 구성된 단축형을 실시할 경우 이 보정 방법이 주는 이득이 더 크다. 또한 Ehrler 등(2019)의 연구에서 제안한 보정회귀 방식은 공식을 도출한 집단과 동질한 집단의 경우만 타당도를 높이고, 동질하지 않은 집단에 적용할 경우 RTM 현상이 반복되는 반면 본 연구에서 제안하는 RTM 보정 방식은 다른 집단에 적용할 때도 비슷한 타당도를 보인다. 종합하자면, 실무적인 관점에서 불가피하게 정보 손실로 인한 오차를 감수하고서라도 단축형 지능검사를 사용해야 할 때가 있는데 이 경우 본 연구가 제안하는 보정 방법이 오차를 줄이는 효과적인 방법이라 볼 수 있다.

본 연구의 시사점과 의의는 다음과 같다. 첫째, 본 연구에서는 회귀추정 방식을 채택한

단축형 지능검사에서 지능이 높거나 낮은 양쪽 꼬리값($\pm 1SD$)에 속하는 대상자에게 실시할 경우 발생하는 체계적 오차, 평균보다 높은 경우 과소추정, 평균보다 낮은 경우 과대추정되는 문제를 해결하였다. 단축형 지능검사가 지적 능력이 낮은 사람들을 선별해야하는 환경에서 선별검사, 신경인지의 손상이 있을 수 있는 뇌손상 환자, 조현병 환자 등 임상 집단에 주로 사용된다는 것을 감안하면 낮은 지능 집단에서 추정IQ가 왜곡되는 것은 치명적인 단점이다. 따라서 RTM 보정 방식을 제안한 것은 임상 실무적으로도 의의가 크다고 할 수 있다. 본 연구의 실질적 활용을 위해 추정IQ에 대응하는 RTM amount와 보정된 추정IQ를 부록에 제시하였으며 RTM 보정 계산기 (<https://trello.com/c/DoWEwrKI>)를 이용하면 간편하게 사용할 수 있다. 둘째, RTM 보정 방식이 다른 추정 방식에 비해 일반화 가능성이 높음을 확인하였다. 기존 보정회귀 방식은 특정 집단의 오차를 상당 폭 줄여주지만 그 보정회귀 공식을 도출한 집단에만 정확도가 높아 해당 집단에서만 유효하게 사용할 수 있다. 반면 RTM 보정 방식은 적용할 대상의 IQ 수준과 무관하게 사용할 수 있다. 셋째, 전통적인 단축형 지능검사 타당도 검증 방법의 문제를 본 연구 결과를 토대로 논증하였고, 더 나은 대안을 제시하였다. 넷째, 단축형 지능검사 추정IQ의 오차 크기를 체계적으로 검토하는 타당도 검증 방식을 통해 선행연구들의 모순된 결과에 대한 통합적인 설명을 제공하였다.

상기와 같은 의의에도 불구하고 본 연구의 한계는 분명하다. 첫째, 본 연구에서는 지적장애 진단 일치도를 살펴보지 않았다. 다만 대부분의 선행연구에서 진단 일치도가 낮았으며, 단축형 지능검사는 진단 목적이라기보다

는 선별 목적이기 때문에 진단 일치도를 고려하는 것이 필수적이라 보기는 어렵다고 생각된다. 둘째, 어떤 소검사를 몇 개 포함할 것인가와 같은 단축형 지능검사 자체의 타당도에 대해서는 본 연구에서 다루지 않았다. 따라서 어떤 단축형 지능검사를 어떤 방식으로 구성하는 것이 최선의 타당도를 보장하는지는 여전히 의문이며 추후 연구에서 심도 있게 다뤄볼 필요가 있다. 셋째, RTM 보정 방식을 다양한 집단과 상황에 모두 일반화할 수 있는지는 본 연구에 포함된 타당화 집단만으로는 확인하기 어렵다. 그러므로 일반화 가능성을 높이기 위해 추후 연구에서는 표준화 자료와 같은 넓은 범위의 IQ 분포를 지닌 대규모의 자료나 평균 이상의 지능을 지닌 집단 등 다양한 표본에 대한 교차 타당화가 이루어져야 할 것으로 보이며 시뮬레이션 연구를 통한 검증도 필요해 보인다. 끝으로 NRMSE, MAPE를 비교하여 RTM 보정 공식이 다른 추정 방식에 비해 정확도가 높음을 제안하였고, 합리적인 기준(MAPE 10% 미만)을 설정하여 각 추정 방식의 타당도를 확인하였지만 추정IQ와 전체척도IQ 간의 동질성을 통계적 가설검증으로 확인하지는 못했다. 그러므로 보정된 추정IQ 일지라도 전체척도IQ와 동일한 의미를 부여해서는 안 될 것이다. 주지하다시피 지적능력을 정확하게 평가하려면 지능검사의 전체척도를 모두 실시하는 것이 최선이다. 단축형으로 추정된 IQ는 어디까지나 추정치일 뿐이므로 편의적 목적으로 전체척도 지능검사를 대체하거나, 진단이나 법적인 의사결정과 같은 중대한 목적으로 사용하는 것은 지양해야 할 것이다.

참고문헌

- Anastasi, A. U., & Urbina, S. S.(1997). *Psychological testing*. International Edition, 7th.
- Barnett, A. G., Van Der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: what it is and how to deal with it. *International journal of epidemiology*, 34(1), 215-220.
<https://doi.org/10.1093/ije/dyh299>
- Basso, M. R., Bornstein, R. A., Roper, B. L. & McCoy, V. L. (2000). Limited accuracy of premorbid intelligence estimators: a demonstration of regression to the mean. *The Clinical Neuropsychologist*. 14, 325-340.
[https://doi.org/10.1076/1385-4046\(200008\)14:3;1-P;FT325](https://doi.org/10.1076/1385-4046(200008)14:3;1-P;FT325)
- Bland, J. M., & Altman, D. G. (1994). Statistics notes: some examples of regression towards the mean. *British Medical Journal*, 309(6957), 780.
<https://doi.org/10.1136/bmj.309.6957.780>
- Bland, M. (2013, February 28). Regression towards the mean or Why was Terminator III such a disappointment? Retrieved form
<https://www-users.york.ac.uk/~mb55/talks/regmean.htm>
- Blyler, C. R., Gold, J. M., Iannone, V. N., & Buchanan, R. W. (2000). Short form of the WAIS-III for use with patients with schizophrenia. *Schizophrenia research*, 46(2-3), 209-215.
[https://doi.org/10.1016/S0920-9964\(00\)00017-7](https://doi.org/10.1016/S0920-9964(00)00017-7)
- Brooks, B. L., & Weaver, L. E. (2005). Concurrent validity of WAIS-III short forms in a geriatric sample with suspected dementia: verbal, performance and full scale IQ scores. *Archives of Clinical Neuropsychology*, 20(8), 1043-1051.
<https://doi.org/10.1016/j.acn.2005.06.005>
- Bulzacka, E., Meyers, J. E., Boyer, L., Le Gloahec, T., Fond, G., Szöke, A., Leboyer M., & Schürhoff, F. (2016). WAIS-IV seven-subtest short form: validity and clinical use in schizophrenia. *Archives of Clinical Neuropsychology*, 31(8), 915-925.
<https://doi.org/10.1093/arclin/acw063>
- Choe, A. Y., Hwang, S. T., Kim, J. H., Park, K. B., Chey, J., & Hong, S. H. (2014). Validity of the K-WAIS-IV short forms. *Korean Journal of Clinical Psychology*, 33(2), 413-428.
<https://doi.org/10.15842/kjcp.2014.33.2.011>
- Chinn S., & Heller R. F. (1981). Some further results concerning regression to the mean. *American Journal of Epidemiology*, 114, 902-905.
<https://doi.org/10.1093/oxfordjournals.aje.a113260>
- Clara, I. P., & Huynh, C. L. (2003). Four short-form linear equation estimates of Wechsler Adult Intelligence Scale III IQs in an elderly sample. *Measurement & Evaluation in Counseling & Development*, 35(4), 251-262.
<https://doi.org/10.1080/07481756.2003.12069071>
- Crawford, J. R., Allan, K. M., & Jack, A. M. (1992). Short-forms of the UK WAIS-R: Regression equations and their predictive validity in a general population sample. *British Journal of Clinical Psychology*, 31(2), 191-202.
<https://doi.org/10.1111/j.2044-8260.1992.tb00983.x>

- Crawford, J. R., Anderson, V., Rankin, P. M., & MacDonald, J. (2010). An index-based short-form of the wisc-iv with accompanying analysis of the reliability and abnormality of differences. *British Journal of Clinical Psychology*, 49(2), 235-258.
<https://doi.org/10.1348/014466509X455470>
- Davis, C. E. (1976) The effect of regression to the mean in epidemiologic and clinical studies. *American Journal of Epidemiology*, 104(5), 493-498
<https://doi.org/10.1093/oxfordjournals.aje.a112321>
- Doppelt, J. E. (1956). Estimating the full scale score on the Wechsler Adult Intelligence Scale from scores on four subjects. *Journal of Consulting Psychology*, 20(1), 63.
<https://doi.org/10.1037/h0044293>
- Ehrler, M., Latal, B., Polentarutti, S., von Rhein, M., Held, L., & Wehrle, F. M. (2019). Pitfalls of using IQ short forms in neurodevelopmental disorders: a study in patients with congenital heart disease. *Pediatric research*, 1-7.
<https://doi.org/10.1038/s41390-019-0667-2>
- Girard, T. A., Axelrod, B. N., Patel, R., & Crawford, J. R. (2014). Wechsler Adult Intelligence Scale-IV dyads for estimating global intelligence. *Assessment*, 22(4), 441-448.
<https://doi.org/10.1177/1073191114551551>
- Haynes, J. P., & Atkinson, D. (1983). Validity of two WPPSI short forms in outpatient clinic settings. *Journal of Clinical Psychology*, 39(6), 961-964.
[https://doi.org/10.1002/1097-4679\(198311\)39:6<961::AID-JCLP2270390623>3.0.CO;2-O](https://doi.org/10.1002/1097-4679(198311)39:6<961::AID-JCLP2270390623>3.0.CO;2-O)
- Hoffman, R. G., & Nelson, K. S. (1988). Cross-validation of six short forms of the WAIS-R in a healthy geriatric sample. *Journal of Clinical Psychology*, 44(6), 952-957.
[https://doi.org/10.1002/1097-4679\(198811\)44:6<952::AID-JCLP2270440617>3.0.CO;2-Y](https://doi.org/10.1002/1097-4679(198811)44:6<952::AID-JCLP2270440617>3.0.CO;2-Y)
- Hwang, K. S., & Oh, S. W. (2017). Validity of the K-WISC-IV Short Forms. *Korean Journal of Clinical Psychology*, 36(3), 381-390.
<http://doi.org/10.15842/KJCP.2017.36.3.008>
- Hwang, S. T., Kim, J. H., Park, K. B., Chey, J., & Hong, S. H. (2012). *K-WAIS-IV: technical and interpretive manual*. Korea psychology.
- Jeong, D. H., Koh, S. H., Kim, J. H., Kang, S. G., Bae, S. M., & Kang, J. M. (2017). Validation of K-WAIS-IV Short Forms among Patients with Traumatic Brain Injury. *Clinical Psychology in Korea: Research and Practice*, 3(1), 23-45.
<http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07193416>
- Jun, Y. S., Hwang, S. T., & Lee S. H. (2008). Validity of the Korean Wechsler Intelligence Scale for Children-III Short forms. *Korean Journal of Clinical Psychology*, 27(1), 277-290.
<http://doi.org/10.15842/kjcp.2008.27.1.016>
- Jung, E. J., & Baek, J. M., (2019). Validity of the K-WAIS-IV Short Forms: Focused on Clinical Utility. *Clinical Psychology in Korea: Research and Practice*, 5(2), 213-231.
<https://doi.org/10.15842/cprp.2019.5.2.213>
- Kaufman, J. C., & Kaufman, A. S. (2001). Time for the changing of the guard: A farewell to short forms of intelligence tests. *Journal of Psychoeducational Assessment*, 19(3), 245-267.

- <https://doi.org/10.1177/073428290101900305>
Lewis, C. D. (1982). *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth-Heinemann.
- Lim, Y. R., Lee, W. K., Lee, W. H., & Park, J. W. (2000). The Study on the Accuracy and Validity of Korean Wechsler Intelligence Scale short forms: A Comparison of the WARD7 subtest vs Doppelt subtest. *Korean Journal of Clinical Psychology, 19*(3), 563-574.
<http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06370902>
- Margolis, R. B., Taylor, J. M., & Greenleaf, C. L. (1986). A cross-validation of two short forms of the WAIS-R in a geriatric sample suspected of dementia. *Journal of clinical psychology, 42*(1), 145-146.
[https://doi.org/10.1002/1097-4679\(198601\)42:1<145::AID-JCLP2270420124>3.0.CO;2-V](https://doi.org/10.1002/1097-4679(198601)42:1<145::AID-JCLP2270420124>3.0.CO;2-V)
- Meyers, J. E., Zellinger, M. M., Kockler, T., Wagner, M., & Miller, R. M. (2013). A validated seven-subtest short form for the WAIS-IV. *Applied Neuropsychology: Adult, 20*(4), 249-256.
<https://doi.org/10.1080/09084282.2012.710180>
- Resnick, R. J. (1977). An abbreviated form of the WISC-R: Is it valid?. *Psychology in the Schools, 14*(4), 426-429.
[https://doi.org/10.1002/1520-6807\(197710\)14:4<426::AID-PITS2310140408>3.0.CO;2-E](https://doi.org/10.1002/1520-6807(197710)14:4<426::AID-PITS2310140408>3.0.CO;2-E)
- Resnick, R. J., & Entin, A. D. (1971). Is an abbreviated form of the WISC valid for Afro-Americans children?. *Journal of Consulting and Clinical Psychology, 36*(1), 97-99.
<https://doi.org/10.1037/h0030475>
- Ryan, J. J., Kreiner, D. S., Gontkovsky, S. T., & Glass Umfleet, L. (2015). Classification accuracy of sequentially administered WAIS-IV short forms. *Applied Neuropsychology: Adult, 22*(6), 409-414.
<https://doi.org/10.1080/23279095.2014.953677>
<https://doi.org/10.1080/23279095.2014.953677>
- Sattler, J. M. (2001). *Assessment of children: Cognitive foundations*. CA: Author.
- Satz, P., & Mogel, S. (1962). An abbreviation of the WAIS for clinical use. *Journal of clinical psychology, 18*(1), 77-79.
[https://doi.org/10.1002/1097-4679\(196201\)18:1<77::AID-JCLP2270180124>3.0.CO;2-R](https://doi.org/10.1002/1097-4679(196201)18:1<77::AID-JCLP2270180124>3.0.CO;2-R)
- Silverstein, A. B. (1971). A corrected formula for assessing the validity of WAIS, WISC, and WPPSI short forms. *Journal of Consulting and Clinical Psychology, 45*, 1162-1163.
[https://doi.org/10.1002/1097-4679\(197104\)27:2<212::AID-JCLP2270270216>3.0.CO;2-8](https://doi.org/10.1002/1097-4679(197104)27:2<212::AID-JCLP2270270216>3.0.CO;2-8)
- Silverstein, A. B. (1982). Two-and four-subtest short forms of the Wechsler Adult Intelligence Scale-Revised. *Journal of Consulting and Clinical Psychology, 50*(3), 415.
<https://doi.org/10.1037/0022-006X.50.3.415>
- So, L. N., & Choi, J. O. (2019) The Validity of K-WAIS-IV Short Form: Clinical Sample. *Clinical Psychology in Korea: Research and Practice, 5*(3), 255-272.
<https://doi.org/10.15842/cprp.2019.5.3.255>
- Tellegen, A., & Briggs, P. F. (1967). Old wine in new skins: Grouping Wechsler subtests into new scales. *Journal of consulting psychology, 31*(5), 499.

- <https://doi.org/10.1037/h0024963>
- Ward, L. C. (1990). Prediction of verbal, performance, and full scale IQs from seven subtests of the WAIS-R. *Journal of Clinical Psychology*, 46(4), 436-440.
[https://doi.org/10.1002/1097-4679\(199007\)46:4<436::AID-JCLP2270460411>3.0.CO;2-M](https://doi.org/10.1002/1097-4679(199007)46:4<436::AID-JCLP2270460411>3.0.CO;2-M)
- Yeo, H. H., Jung, J. H., Kim C. H., Seo, M. J., Cho, Y. I., & Chang E. J. (2019). The purpose of this study was to improve the existing short-form (SF) of the K-WAIS-IV to be used in the Korean military draft examination. *The Korean Society for Applied Developmental Psychology*, 9(1), 67-84.
<https://doi.org/10.22839/adp.2020.9.1.67>
- 1차원고접수 : 2020. 09. 04.
2차원고접수 : 2021. 01. 07.
3차원고접수 : 2021. 04. 06.
최종게재결정 : 2021. 05. 17.

The systematic errors of K-WAIS-IV Short Forms and new correction proposal

Hwan-Hong Yeo¹⁾

Yoonhyoung Lee²⁾

¹⁾Chung-nam Branch, Military Manpower Administer

²⁾Department of Psychology, Yeungnam University

Using regression equation for short forms of intelligence scales tends to cause over-estimation of the IQ especially for those who have lower intelligence. Such a systematic error in IQ estimation threatens the validity of intelligence scale short form. The current study aims to resolve the problem by proposing new way of correction based on RTM amount formula by Davis(1976). To do so, this study compared the accuracies of proration method, linear equating method, regression formula method, regression equation correction by Ehrler et al.(2019) with our correction proposal. Two homogeneous low intelligent groups and a heterogeneous clinical group were tested. As results, the proposed correction method showed better accuracy over regression formula method, better validity over proration method and linear equating method, and more generally applicable over Ehrler et al.(2019)'s correction. The proposed correction method successfully reduced the systematic error of over-estimation for lower intelligence.

Key words : K-WAIS-IV, short-form, estimated IQ, regression toward mean, systematic error correction

부 록

RTM 보정 전, 후 추정IQ

2개 소검사 단축형			4개 소검사 단축형		
추정IQ	RTM amount	RTM 보정 추정IQ	추정IQ	RTM amount	RTM 보정 추정IQ
60	20	40	53	13	40
61	20	41	54	12	42
62	20	42	55	12	43
63	19	44	56	12	44
64	19	45	57	12	45
65	18	47	58	12	46
66	18	48	59	11	48
67	17	50	60	11	49
68	17	51	61	11	50
69	17	52	62	11	51
70	16	54	63	10	53
71	16	55	64	10	54
72	15	57	65	10	55
73	15	58	66	10	56
74	14	60	67	9	58
75	14	61	68	9	59
76	14	62	69	9	60
77	13	64	70	9	61
78	13	65	71	9	62
79	12	67	72	8	64
80	12	68	73	8	65
81	12	69	74	8	66
82	11	71	75	8	67
83	11	72	76	7	69
84	10	74	77	7	70
85	10	75	78	7	71
86	10	76	79	7	72
87	9	78	80	7	73
88	9	79	81	6	75
89	9	80	82	6	76
90	8	82	83	6	77
91	8	83	84	6	78
92	7	85	85	6	79
			86	5	81
			87	5	82
			88	5	83
			89	5	84
95% CI 표준오차(±1.96RMSE)					
±30.75		±13.84		±16.07	
				±10.86	