*Georgia State University*
*Project (EPRS 9360)*
*5.1.2008*

# Didactical Derivation of DFIT Indices

**Sarah J. Kim**
*Georgia State University*

Raju, N.S., Van der Linden, W. J., & Fleer, P. F. (1995) IRT-Based Internal Measures of Differential Functioning of Items and Tests. Applied Psychological Measurement, 19, 353-368

## *Differential Test Functioning*

Let $P_i(\theta_s)$ represent the probability of success for examinee *s* with trait level $\theta$ on item *i*. $P_i$ can be represented either by a one-, two-, or three- parameter logistic model or the normal ogive model (Lord, 1980, pp.12-13), with item parameter a (discrimination), b(difficulty), and c (pseudoguessing). Let the test consist of *n* items and have one set of item parameters for each of two groups-the reference (R) group (typically the majority group) and the focal (F) group (typically the minority group). Also assume that the two sets of item parameters are on a common scale. $P_{iR}(\theta)$ represents the probability of success on item *i* at a given $\theta$ level for examinee *s* if examinee *s* is a member of the reference group, $P_{iF}(\theta)$ represents the probability of success on the same item for examinee *s* if examinee *s* is a member of the focal group. If an item functions differently in the two groups, $P_{iR}$ and $P_{iF}$ will be different for some examinees.

(1) Within IRT, an examinee's expected proportion correct (EPC, sometimes referred to as the "true" score)

$$T_s = \sum_{i=1}^{n} P_i(\theta)$$

(2)

$$DTF = E\left(T_{sF} - T_{sR}\right)^2,$$

where each examinee has two EPCs –one as a member of the focal group and the other as a member of the reference group.

(3)

$$DTF = E_F\left(T_{sF} - T_{sR}\right)^2,$$

where the expectation (∈) can be taken over the reference group or the focal group.

(4) $DTF = E_F\left(D_s^{\,2}\right) = \int D_s^{\,2} f_F(\theta)\, d\theta = \sigma_D^{\,2} + \left(\mu_{TF} - \mu_{TR}\right)^2 = \sigma_D^{\,2} + \mu_D^{\,2},$

where $f_F(\theta)$ is the density function of $\theta$ in the focal group, and $\mu_{TF}$ and $\mu_{TR}$ represent the mean EPC of examinees in the focal and reference groups, respectively.

Proof:

$\sigma_D^{\,2} + \mu_D^{\,2}$

* Hint 1. Here, $\sigma_D^{\,2} + \mu_D^{\,2} = E_F\left[D^2\right]$ is derived from the definition of variance and expectation.

$= E_F\left[\left(D - \mu_D\right)^2\right] + \mu_D^{\,2}$

* Hint 2. Use the formula of $Var(X) = E(X - \mu)^2$ and see the guide of mathematical statistical guide for understanding further.

$$= E_F\left[D^2 - 2D\cdot\mu_D + \mu_D{}^2\right] + \mu_D{}^2$$

* Hint 3. In calculus, $(a+b)^2 = a^2 + 2ab + b^2$. Apply this formula to $\left(D-\mu_D\right)^2$

$$= E_F\left[D^2\right] - E_F\left[2\mu_D D\right] + E_F\left[\mu_D{}^2\right] + \mu_D{}^2$$

* Hint 4. Use the property of expectation value, which is $E(ax+b) = aE(x)+b$

$$= E_F\left[D^2\right] - 2\mu_D E_F\left[D\right] + \mu_D{}^2 + \mu_D{}^2$$

* Hint 5. $E_F\left[2\cdot\mu_D\cdot D\right] = 2\mu_D\cdot E(D)$, because $2\mu_D$ is constant value. According to the property of expectation, expectation of constant value is constant, such as $E(a) = a$

$$= E_F\left[D^2\right] - 2\mu_D E_F\left[D\right] + \mu_D{}^2 + \mu_D{}^2$$

$$= E_F\left[D^2\right] - \mu_D{}^2 + \mu_D{}^2$$

$$= E_F\left[D^2\right]$$

***Differential Item Functioning***

***A compensatory DIF index (Based on Equation 1, and 2 can be rewrited as)***

(5) $\text{DTF} = E\left[\left(\sum_{i=1}^{n} d_{is}\right)^2\right]$,

where $d_{is} = P_{iF}(\theta_s) - P_{iR}(\theta_s)$

(6) $\text{DTF} = \sum_{i=1}^{n}\left[Cov(d_i, D) + \mu_{d_i}\mu_D\right]$,

where $Cov(d_i, D)$ is the covariance between the difference in item probabilities for item $i$ $(d_i)$ and the difference between the two EPCs (D), and $\mu_{d_i}$ and $\mu_D$ are the mean of $d_i$ and $D_s$, respectively.

$$DTF = E\left[\left(\sum_{i=1}^{n} d_{is}\right)^2\right]$$

$$= E\left[\left(\sum_{i=1}^{n}(P_{iF} - P_{iR})\right)^2\right] = E\left[\left\{\sum_{i=1}^{n} P_{iF} - \sum_{i=1}^{n} P_{iR}\right\}^2\right]$$

* Hint 6. $\left[E(\sum a_i)\right]^2 = \left[E(\sum a_i)^2\right]$

For example, Let say $a_1 = 2,\ \ a_2 = 3,\ \ a_3 = 4$.

$\left[E(\sum a_i)\right]^2 = \left[E(2+3+4)\right]^2 = \left[E(9)\right]^2 = [9]^2 = 81$,

$\left[E(\sum a_i)^2\right] = \left[E(2+3+4)^2\right] = E(9)^2 = E(81) = 81$.

$$= E\left[\left(\sum_{i=1}^{n} P_{iF} - \sum_{i=1}^{n} P_{iR}\right)\cdot\left(\sum_{i=1}^{n} P_{iF} - \sum_{i=1}^{n} P_{iR}\right)\right]$$

*Hint 7. In calculus,

$(a \pm b)^2 = (a \pm b)\cdot(a \pm b)$.

$$= E\left[\sum_{i=1}^{n}(P_{iF} - P_{iR})\cdot\left(\sum_{i=1}^{n} P_{iF} - \sum_{i=1}^{n} P_{iR}\right)\right]$$

* Hint 8. Using Hint 6,

$\sum(P_F) - \sum(P_R) = \sum(P_F - P_R)$

$$= E\left[\sum_{i=1}^{n} d_{is}\cdot(T_{sF} - T_{sR})\right]$$

$$= \sum_{i=1}^{n}\left[E(d_{is}\cdot D)\right]$$

*Hint 9. Using covariance formula, which is

$Cov(X,Y) = E[XY] - E[X]E[Y]$

$$= \sum_{i=1}^{n}\left[E(d_{is}\cdot D) - E(d_{is})E(D) + E(d_{is})E(D)\right]$$

* Hint 10. Manipulate the equation by adding and deleting the term of $E(d_{is})E(D)$, in order to inducing covariance formula.

$$= \sum_{i=1}^{n}\left[Cov(d_{is}\cdot D) + E(d_{is})E(D)\right]$$

* Hint 11. According to the formula at Hint 10, $E(d_{is}\cdot D) - E(d_{is})E(D) = Cov(d_{is}\cdot D)$

$$= \sum_{i=1}^{n}\left[Cov(d_{is}\cdot D) + \mu_{d_i}\mu_D\right],$$

where $D_s = T_{sF} - T_{sR}$, and $d_{iS} = P_{iF}(\theta_s) - P_{iR}(\theta_s)$

(7) $CDIF_i = E\left(d_i \cdot D\right)$ * Hint 13. It is the same procedure as (6) equation. See Hint 9-11.

$$\left[E(d_{is} \cdot D)\right] = E(d_{is} \cdot D) - E\left(d_{is}\right)E\left(D\right) + E\left(d_{is}\right)E\left(D\right)$$ * Hint 14. See Hint 10.

$$= Cov(d_{is} \cdot D) + E\left(d_{is}\right)E\left(D\right)$$

$$= Cov(d_{is} \cdot D) + \mu_{d_i}\mu_D,$$

where $D_s = T_{sF} - T_{sR}$, and $d_{iS} = P_{iF}(\theta_s) - P_{iR}(\theta_s)$.

(8) $DTF = \sum_{i=1}^{n} CDIF_i$,

which shows that the definition of $CDIF_i$ is additive in the sense that differential functioning at the test level is simply the sum or differential functioning at the item level, and which indicates how much each item's CDIF contributes to DTF. Furthermore, rewriting Equation 5 yields

(9) $DTF = E\left[\sum_{i=1}^{n}\left(P_{iF} - P_{iR}\right)^2\right] = E\left[\left(P_{1F} - P_{iR}\right) + \left(P_{2F} - P_{2R}\right) + \cdots + \left(P_{nF} - P_{nR}\right)\right]^2$

***A Noncompensatory DIF index***

If it is assumed that all items in the test other than item I are completely unbiased, then it must be true that $d_j = 0$ at all $i \neq j$. Equation 7 ($CDIF_i = E(d_i \cdot D)$) can be rewritten as

(10) $NCDIF_i = \sigma_{d_i}{}^2 + \mu_{d_i}{}^2$,

which does not include information about bias from other items.

$CDIF_i = E(d_i \cdot D)$ * Hint 15. It is the same procedure as (6) equation. See Hint 9-11.

$$= E(d_i \cdot d_i) = E\left(d_i^{\,2}\right)$$

$$= E\left(d_i^{\,2}\right) - E\left(d_i\right)^2 + E\left(d_i\right)^2$$

$$= \sigma_{d_i}{}^2 - \mu_{d_i}{}^2,$$

where $d_i = D$

If $d_j = 0$ for all $i \neq j$,

$$D_s = \left(T_{sF} - T_{sR}\right)$$

$$= \sum_{i=1}^{n} P_{iF}\left(\theta_s\right) - \sum_{i=1}^{n} P_{iR}\left(\theta_s\right)$$

$$= \left(P_{1F} + P_{2F} + \cdots + P_{nF}\right) - \left(P_{1R} + P_{2R} + \cdots + P_{nR}\right)$$

$$= P_{iF} - P_{iR} = d_{is}$$

(11) $NCDIF_i = \int_{-\infty}^{\infty} \left[P_{iF}(\theta) - P_{iR}(\theta)\right]^2 f_F\left(\theta\right) d\theta$

By letting $f_F\left(\theta\right)$ denote the density function of $\theta$ in the focal group, Equation 10 can be rewritten as $CDIF_i = E(d_i \cdot D) = E(d_i \cdot d_i) = E\left(d_i^{\,2}\right)$

$NCDIF_i = E\left(d_i^{\,2}\right) = \int_{-\infty}^{\infty} d_i^{\,2} f_F\left(\theta\right) d\theta$ * Hint 16. If $X$ is a continuous random variable having a probability density function $f(x)$, $E(X) = \int_{-\infty}^{\infty} xf(x)dx$. See the mathematical statistics guide below.

$$= \int_{-\infty}^{\infty} \left[P_{iF}(\theta) - P_{iR}(\theta)\right]^2 \; f_F\left(\theta\right) d\theta \, ,$$

where $d_{iS} = P_{iF}(\theta_s) - P_{iR}(\theta_s)$.

(12) Equation 11 can be rewritten as

$$NCDIF_i = \int_{-\infty}^{\infty} \left|P_{iF}(\theta) - P_{iR}(\theta)\right|^2 \; f_F\left(\theta\right) d\theta$$

* Hint 17. In calculus, $\left(a \pm b\right)^2$ is the same as $\left|a \pm b\right|^2$. So $\left[P_{iF}(\theta) - P_{iR}(\theta)\right]^2 = \left|P_{iF}(\theta) - P_{iR}(\theta)\right|^2$

(13) According to Cauchy-Schwartz inequality, Equation 12 can be expressed as

$$NCDIF_i \geq \left[\int_{-\infty}^{\infty} \left|P_{iF}\left(\theta_s\right) - P_{iR}\left(\theta_s\right)\right| f_F\left(\theta\right) d\theta\right]^2,$$

*Hint 18. According to Cauchy Schwartz inequality, $\left|\left(x, y\right)\right| \leq |x| \cdot |y|$

where by the definition of Cauchy-Schwartz inequality.

(14)- (18) Equation 14, 15, 16, 17, and 18 are estimates of DTF, CDF, and NCDIF, respectively.

- **Expectation (Ross, 1988):** In probability theory, the expected value of a discrete random variable is the sum of the probability of each possible outcome of the experiment multiplied by the outcome value. Therefore, if *X* is a discrete random variable having a probability mass function *p(x)*, the expectation or the expected value of X, denoted by *E(X)* is defined by

$$E(X) = \sum xp(x)$$

If $X$ is a continuous random variable having a probability density function $f(x)$, then as $f(x)dx \approx P\{x \le X \le x + dx\}$ for $dx$ small, it is reasonable to define the expected value of X by

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

1. **Example:** Assume that one rolls a pair of balanced dice. Let random variables X be the face values of the die.

   Solution: Since $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$, we obtain that

$$E(X) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + \left(\frac{1}{6}\right) = \frac{7}{2}$$

2. **Example (Ross, 1988):** Expectation of a Normal Random variable. When X is normally distributed with parameters $\mu$ and $\sigma^2$, $E(X) = \mu$

   Solution: The density of a normal random variable is $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}$.

$$E(X) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} xe^{-(x-\mu)^2/2\sigma^2} dx$$

Letting $x = (x - \mu) + \mu$ yields

$$E(X)=\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{\infty}(x-\mu)e^{-(x-\mu)^2/2\sigma^2}dx+\mu\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{\infty}e^{-(x-\mu)^2/2\sigma^2}dx$$

Letting $y=x-\mu$ in the first integral

$$E(X)=\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{\infty}ye^{-y^2/2\sigma^2}dy+\mu\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma}}e^{-(x-\mu)^2/2\sigma^2}dx$$

$$E(X)=0+\mu\cdot 1=\mu,$$

Where f(x) is the normal density. So

$$E(X)=\mu$$

- **Variance (Ross, 1988);** If X is a random variable with mean $\mu$, then the variance of X is defined by $Var(X)=E(X-\mu)^2$

$$Var(X)=E(X-\mu)^2$$

$$=E\left[X^2-2\mu X+\mu^2\right]=E\left[X^2\right]-E\left[2\mu X\right]+E\left[X\right]^2$$

$$=E\left[X^2\right]-2\mu E\left[X\right]+\mu^2=E\left[X^2\right]-2\mu\mu+\mu^2$$

$$=E\left[X^2\right]-\mu^2$$

3. **Example (Ross, 1988):** Variance of a Normal random Variable. When x is normal random variable with parameters $\mu$ and $\sigma^2$,

$$Var(X)=\sigma^2$$

Solution: According to Example 2,

$$E(X)=\mu\text{, and }\ Var(X)=E(X-\mu)^2$$

$$E(X-\mu)^2 = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx$$

Letting $y = (x-\mu)/\sigma$

$$Var(X) = \frac{\sigma^2}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \left[ -ye^{-y^2/2} \Big| + \int_{-\infty}^{\infty} e^{-y^2/2} dy \right] \text{ by integration by parts}$$

$$= \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = \sigma^2 \cdot 1 = \sigma^2,$$

where $\int_{-\infty}^{\infty} f(x)dx = 1$

- **Covariance (Ross, 1988).:** The covariance of any two random variables X and Y, defined by

$$Cov(X,Y) = E\left[\left(X - E(X) \cdot (Y - E(Y)\right)\right]$$

$$= E\left[XY - E(X)Y - XE(Y) + E(X)E(Y)\right]$$

$$= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y]$$

$$= E[XY] - E[X]E[Y]$$

## Reference

Ross, S.: *A First Course in Probability Third Edition*. New York: Maxwell Macmillan, 1989, 422pp.