PSYCHOMETRIKA—VOL. 84, NO. 4, 1129–1151
DECEMBER 2019
https://doi.org/10.1007/s11336-019-09674-5

THE ASYMPTOTIC DISTRIBUTION OF AVERAGE TEST OVERLAP RATE IN COMPUTERIZED ADAPTIVE TESTING

EDISON M. CHOE

GRADUATE MANAGEMENT ADMISSION COUNCIL™ (GMAC™)

HUA-HUA CHANG

PURDUE UNIVERSITY

The average test overlap rate is often computed and reported as a measure of test security risk or item pool utilization of a computerized adaptive test (CAT). Despite the prevalent use of this sample statistic in both literature and operations, its sampling distribution has never been known nor studied in earnest. In response, a proof is presented for the asymptotic distribution of a linear transformation of the average test overlap rate in fixed-length CAT. The theoretical results enable the estimation of standard error and construction of confidence intervals. Moreover, a practical simulation study demonstrates the statistical comparison of average test overlap rates between two CAT designs with different exposure control methods.

Key words: test overlap, item exposure, test security, pool utilization, computerized adaptive testing, asymptotic theory.

1. Introduction to Computerized Adaptive Testing

Computerized adaptive testing (CAT) is generally well regarded as an efficient vehicle for measuring a person's latent trait(s) of interest. The primary reason is that the engine of a proper CAT is an item selection algorithm that seeks to deliver a tailored exam for every participant. This is accomplished by adapting to an individual test-taker's performance in real time, which necessitates a stochastic model of responses as a function of latent trait(s) and item characteristics. Item response theory (IRT) provides the requisite framework, within which the three-parameter logistic model (3PLM; Lord & Novick, 1968) is commonly employed for measuring univariate ability with dichotomously scored items:

$$P_j(\theta) := P(X_j = 1|\theta) = c_j + \frac{1 - c_j}{1 + e^{-a_j(\theta - b_j)}}. \tag{1}$$

In the context of cognitive assessment, θ is the latent ability parameter and X_j is a binary random variable that maps correct and incorrect responses on item j to 1 and 0, respectively. Therefore, $P_j(\theta)$ models the probability of a correct answer on item j as a function of ability, where a_j, b_j, and c_j represent the item's discrimination, difficulty, and pseudo-guessing parameters, respectively. In operational practice, new items are typically pretested on an initial sample of examinees and calibrated using an estimation technique such as the marginal maximum likelihood (MML; Baker & Kim 2004). During estimation, θ and b are always placed on the same fixed scale to establish a direct correspondence between person ability and item difficulty. The item parameter estimates are then treated as known values in subsequent usage.

Correspondence should be made to Edison M. Choe, Graduate Management Admission Council™ (GMAC™), 11921 Freedom Drive, Suite 300, Reston, VA 20190, USA. Email: echoe@gmac.com

CAT algorithms are commonly based on selecting the next item with the maximum Fisher information given θ (MFI; Lord, 1980). For 3PLM, the item information function is derived as

$$I_j(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\log L(\theta|x_j)\right) = a_j^2\left(\frac{1-P_j(\theta)}{P_j(\theta)}\right)\left(\frac{P_j(\theta)-c_j}{1-c_j}\right)^2, \tag{2}$$

in which $L(\theta|x_j)$ is the likelihood function of θ given an observed response x_j:

$$L(\theta|x_j) = P_j(\theta)^{x_j}[1-P_j(\theta)]^{1-x_j}. \tag{3}$$

Of course, an examinee's true ability is unknown in practice, which is precisely the reason for measuring it. Hence, information is estimated using the current interim estimate of ability, which is often computed using maximum likelihood as

$$\hat{\theta} = \arg\max_{\theta} L(\theta|\mathbf{x}) = \arg\max_{\theta} \prod_{j=1}^{k} P_j(\theta)^{x_j}[1-P_j(\theta)]^{1-x_j}. \tag{4}$$

Here, $\mathbf{x} = \{x_1, x_2, \ldots, x_k\}$ is the set of observed responses to a corresponding set of k items that have been administered thus far. The standard error of estimation can then be approximated as the inverse of the square root of the estimated information accumulated across the k items:

$$\mathrm{SE}(\theta) \approx \left[\sum_{j=1}^{k} I_j(\hat{\theta})\right]^{-1/2}. \tag{5}$$

Note that maximum likelihood estimation (MLE) technically depends on the assumption that every new response is independent of the previous ones strictly conditional on θ only (commonly called "local independence" in IRT literature). However, this is violated by design in CAT since the modus operandi of any adaptive algorithm is to select subsequent items based on the previous responses (Mislevy & Chang, 2000). Nevertheless, Chang and Ying (2009) used martingale theory to prove that, under mild regularity conditions and when MFI is used to select items from an unlimited pool, $\hat{\theta}$ still converges in distribution as follows:

$$\hat{\theta} \xrightarrow{d} \mathcal{N}\left(\theta, \frac{1}{I^{(k)}(\theta)}\right) \quad \text{as} \quad k \to \infty, \tag{6}$$

where $I^{(k)}$ is the cumulative information function for k items (see also Chang 2015). Therefore, under ideal circumstances, CAT running a pure MFI algorithm with MLE guarantees a consistent and efficient estimate of θ.

In practice with a finite pool, however, an inevitable trade-off for such optimal measurement is a highly disproportionate usage of items. This is unsurprising since the algorithm would tend to prioritize only a handful of items with the most information, which are usually those with the largest a values (Chang & Ying, 1999; Chang, Qian, & Ying, 2001; Hau & Chang, 2001). Intuitively, overexposed items generally pose a greater security risk of unauthorized disclosure, potentially leading to compromise of test integrity and score validity. Furthermore, overexposure entails underexposure of other items, indicating suboptimal management of costly assets. To make matters worse, greater imbalance in item exposure directly results in larger test overlap between examinees (Chen, Ankenmann, & Spray, 2003), which increases the risk of collusion between test takers. Consequently, especially for high-stakes testing, MFI is virtually always restrained with a mechanism that controls item exposure (Way, 1998).

In light of security threats and financial implications of a poorly utilized item pool, a more rigorous understanding of item exposure and test overlap is in order. The material is presented within a mathematical framework and organized into the following sections: defining the concepts of item exposure and test overlap, deriving the baseline distribution of average test overlap rate under random item selection, deriving the focal distribution of average test overlap rate under non-random item selection, and applying the results in an illustrative CAT simulation comparing two exposure control methods. The discourse is limited to fixed-length CAT as the intractable complications of variable-length CAT exceed the intended scope of the present study.

2. Definitions of Item Exposure and Test Overlap

Consider a CAT window consisting of n examinees ($i = 1, \dots, n$) and a pool of m items ($j = 1, \dots, m$). For each examinee i, define $\mathbf{U}_i$ to be an m-vector of binary indicators of whether the jth item was administered:

$$\mathbf{U}_i = [U_{i1} \ \cdots \ U_{im}]'. \tag{7}$$

In other words, $U_{ij} = 1$ if examinee i receives item j and $U_{ij} = 0$ otherwise. Thus, the test length for every examinee can be expressed as an n-vector $\mathbf{L}$ consisting of squared lengths of each $\mathbf{U}_i$, or more intuitively, the sums of each U_{ij} across all m items (since $U_{ij}^2 = U_{ij}$):

$$\mathbf{L} = \left[||\mathbf{U}_1||^2 \ \cdots \ ||\mathbf{U}_n||^2\right]' = \left[L_1 = \sum_{j=1}^{m} U_{1j} \ \cdots \ L_n = \sum_{j=1}^{m} U_{nj}\right]'. \tag{8}$$

Likewise, the exposure count for every item can be expressed as an m-vector $\mathbf{V}$ consisting of the sums of each U_{ij} across all n examinees:

$$\mathbf{V} = \sum_{i=1}^{n} \mathbf{U}_i = \left[V_1 = \sum_{i=1}^{n} U_{i1} \ \cdots \ V_m = \sum_{i=1}^{n} U_{im}\right]'. \tag{9}$$

The total item exposure count N can then be computed as

$$N = \sum_{j=1}^{m} V_j = \sum_{i=1}^{n} L_i. \tag{10}$$

Consequently, the vector of item exposure rates, which can be interpreted as the mean vector of U_{ij}'s across examinees, is given as

$$\bar{\mathbf{U}} = \mathbf{V}/n = \left[\bar{U}_1 = V_1/n \ \cdots \ \bar{U}_m = V_m/n\right]'. \tag{11}$$

TABLE 1.
Item exposures, $U_{ij} = 1$ or 0, for CAT window with $i = 1, \ldots, n$ examinees and $j = 1, \ldots, m$ items.

	1	$\cdots$	j	$\cdots$	m	$\sum_{j=1}^{m} U_{ij}$
$\mathbf{U}_1$	U_{11}	$\cdots$	U_{1j}	$\cdots$	U_{1m}	L_1
$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$	$\vdots$
$\mathbf{U}_i$	$\vdots$	$\cdots$	U_{ij}	$\cdots$	$\vdots$	L_i
$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$
$\mathbf{U}_n$	U_{n1}	$\cdots$	U_{nj}	$\cdots$	U_{nm}	L_m
$\sum_{i=1}^{n} \mathbf{U}_i = \mathbf{V}$	V_1	$\cdots$	V_j	$\cdots$	V_m	$\sum_{i=1}^{n}\sum_{j=1}^{m} U_{ij} = N$
$\bar{\mathbf{U}} = \mathbf{V}/n$	$\bar{U}_1$	$\cdots$	$\bar{U}_j$	$\cdots$	$\bar{U}_m$	$N/n = \bar{L}$

For fixed-length CAT (i.e., $L_i = L \ \forall\ i$), $N = nL$ and $\bar{L} = L$.

Additionally, the sum of $\bar{\mathbf{U}}$'s elements is equal to the average test length across examinees:

$$\bar{\mathbf{U}}'\mathbf{1} = \sum_{j=1}^{m} \bar{U}_j = \frac{1}{n}\sum_{j=1}^{m} V_j = \frac{1}{n}\sum_{i=1}^{n} L_i = \bar{L}, \tag{12}$$

where $\mathbf{1}$ is an m-vector of ones. Note that for fixed-length CAT in which every examinee receives the same number of items L (i.e., $L_i = L \ \forall\ i \in \{1, \ldots, n\}$), $\bar{L} = L$ and $N = nL$. Table 1 summarizes all of this information.

Next, define test overlap to be the number of shared items between a pair of distinct examinees i and i' (i.e., $i \neq i'$), which can be quantified as

$$W_{ii'} = \mathbf{U}_i'\mathbf{U}_{i'} = \sum_{j=1}^{m} U_{ij}U_{i'j}. \tag{13}$$

Then for fixed-length CAT, test overlap rate is given as $R_{ii'} = W_{ii'}/L$, and the average test overlap rate is given by the following lemma.

Lemma 2.1. (Chen et al., 2003) *Let $\bar{U}_j$ be the observed exposure rate of item j after n examinees have each taken a fixed-length CAT consisting of L items drawn from a pool of m items. Then the average test overlap rate can be computed as*

$$\bar{R} = \frac{n}{L(n-1)}\sum_{j=1}^{m} \bar{U}_j^2 - \frac{1}{n-1}. \tag{14}$$

Proof. By definition, the average test overlap rate is the sum of $R_{ii'}$ over all $\binom{n}{2}$ possible pairs of examinees divided by $\binom{n}{2}$:

$$\bar{R} = \frac{\sum_{i=1}^{n-1}\sum_{i'=i+1}^{n} R_{ii'}}{\binom{n}{2}} = \frac{\sum_{i=1}^{n-1}\sum_{i'=i+1}^{n} W_{ii'}}{L\binom{n}{2}}. \tag{15}$$

Note that the sum of $W_{ii'}$ over all pairs of examinees can also be expressed as

$$\sum_{i=1}^{n-1}\sum_{i'=i+1}^{n} W_{ii'} = \sum_{i=1}^{n-1}\sum_{i'=i+1}^{n}\sum_{j=1}^{m} U_{ij}U_{i'j} = \sum_{j=1}^{m}\left\{\sum_{i=1}^{n-1}\sum_{i'=i+1}^{n} U_{ij}U_{i'j}\right\} = \sum_{j=1}^{m}\binom{V_j}{2}. \tag{16}$$

This makes sense since the term in the braces represents the number of examinee pairs both having received item j, which is equivalently but more efficiently computed as $\binom{V_j}{2}$. Therefore, a computation-friendly formulation of average test overlap rate is

$$\bar{R} = \frac{\sum_{j=1}^{m}\binom{V_j}{2}}{L\binom{n}{2}} = \frac{\sum_{j=1}^{m} V_j(V_j-1)}{Ln(n-1)} = \frac{\sum_{j=1}^{m} n\bar{U}_j(n\bar{U}_j-1)}{Ln(n-1)} = \frac{n}{L(n-1)}\sum_{j=1}^{m}\bar{U}_j^2 - \frac{1}{n-1}. \tag{17}$$

□

3. The Baseline Distribution of Average Test Overlap Rate

For the purpose of establishing a benchmark, consider the assumption that a fixed number of items are selected completely at random for every examinee, resulting in a fully balanced utilization of the item pool. This may hold true for certain non-adaptive designs, such as linear on-the-fly testing (LOFT) with uniform distribution of contents. The following results and proofs for the baseline case will lay the groundwork for elucidating the focal case of adaptive item selection presented in the next section.

Lemma 3.1. *Suppose a random sample of n examinees are each* ***randomly*** *administered L items from a pool of m items. Then, $\forall\ i \in \{1,\ldots,n\}$, $\mathbf{U}_{0i}$ are independent and identically distributed (i.i.d.) as m-dimensional Bernoulli with expectation $\mathbf{p}_0$ and covariance $\mathbf{\Sigma}_0$ as follows:*

$$\mathbf{p}_0 = [p \quad p \quad \cdots \quad p]', \quad \mathbf{\Sigma}_0 = \sigma^2\begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix},$$

where $p = L/m$, $\sigma^2 = p(1-p)$, and $\rho = -1/(m-1)$.

Proof. For any given ith examinee, random item selection means a simple random sampling without replacement (SRSWOR) of L items from a pool of m items. [Consult, for example, Cochran's (1977) classic text on sampling techniques for a more general treatment of SRSWOR.] This guarantees an independent and identically uniform distribution of item selection across all examinees. Under such an idealistic scenario, two methods are demonstrated for computing probabilities of item exposures across trials:

- *Method 1*. Define Ω to be the set of every possible ordered sequence of L item exposures for a given trial, which has a cardinality of

$$|\Omega| = {}_{m}P_L = \frac{m!}{(m-L)!}. \tag{18}$$

Then $\omega_s \in \Omega$ represents a unique elementary event $s \in \{1, \ldots, |\Omega|\}$, for example,

$$\omega_9 := \text{ordered sequence of } U_{i7} = 1, U_{i2} = 1, \ldots, \tag{19}$$

whose probability can be computed as

$$P(\omega_9) = \frac{1}{m} \cdot \frac{1}{m-1} \cdots \frac{1}{m-L+1} = \frac{(m-L)!}{m!} = |\Omega|^{-1} \tag{20}$$

without loss of generality. Note that the item selection probability starts at $1/m$ for the first item, then the denominator is subtracted by one at each subsequent selection to normalize the total probability across all remaining items (i.e., to ensure the probabilities sum to one). Since every ω_s is equally probable due to random item selection, the probability of any single instance is $1/|\Omega|$. Next, define $\mathcal{C}_j \subseteq \Omega$ to be the event of the exposure of a particular item j, which consists of every ω_s that includes $U_{ij} = 1$ in any position. The total number of such instances can be counted as the number of possible positions of item j out of L times the number of permutations of the other $L-1$ items out of the remaining pool of $m-1$ items:

$$|\mathcal{C}_j| = [\,{}_{L}P_1][\,{}_{(m-1)}P_{(L-1)}] = L\frac{(m-1)!}{(m-L)!}. \tag{21}$$

Thus, the exposure probability of any given item j can be calculated as

$$p := P(U_{ij} = 1) = \sum_{w_s \in \mathcal{C}_j} P(w_s) = \frac{|\mathcal{C}_j|}{|\Omega|} = \frac{L}{m}. \tag{22}$$

Likewise, define $\mathcal{D}_{jj'} \subseteq \Omega$ to be the event of the exposure of any pair of distinct items j and j' (i.e., $j \neq j'$), which consists of every ω_s that includes both $U_{ij} = 1$ and $U_{ij'} = 1$ in any order. The total number of such instances can be counted as the number of permutations of items j and j' out of L times the number of permutations of the other $L-2$ items out of the remaining pool of $m-2$ items:

$$|\mathcal{D}_{jj'}| = [\,{}_{L}P_2][\,{}_{(m-2)}P_{(L-2)}] = L(L-1)\frac{(m-2)!}{(m-L)!}. \tag{23}$$

Thus, the joint exposure probability of any given pair of items j and j' can be calculated as

$$p^* := P(U_{ij} = 1, U_{ij'} = 1) = \sum_{w_s \in \mathcal{D}_{jj'}} P(w_s) = \frac{|\mathcal{D}_{jj'}|}{|\Omega|} = \frac{L(L-1)}{m(m-1)}. \tag{24}$$

- *Method 2*. Since all *permutations* of L items from a pool of m items are equally likely on any given trial, the problem can be simplified by disregarding item selection order. In other words, there are a total of $\binom{m}{L}$ unique *combinations* of L items from a pool of m items that are all equally probable. The exposure of item j can be perceived as first administering item j and then randomly sampling an additional $L-1$ items from the remaining pool of $m-1$ items without replacement. Thus, there are a total of $\binom{m-1}{L-1}$ unique patterns of $\mathbf{U}_{0i}$ in which a particular U_{ij} is fixed at 1, so the probability of exposure of any item across trials can be calculated as

$$p := P(U_{ij} = 1) = \frac{\binom{m-1}{L-1}}{\binom{m}{L}} = \frac{L}{m}. \tag{25}$$

Similarly, for any two items j and j', there are a total of $\binom{m-2}{L-2}$ unique patterns of $\mathbf{U}_{0i}$ in which both $U_{ij} = 1$ and $U_{ij'} = 1$, so the probability of such an event is

$$p^* := P(U_{ij} = 1, U_{ij'} = 1) = \frac{\binom{m-2}{L-2}}{\binom{m}{L}} = \frac{L(L-1)}{m(m-1)}. \tag{26}$$

Therefore, the expected value and variance of any U_{ij} can be derived as, respectively,

$$E(U_{ij}) = P(U_{ij} = 1) = \frac{L}{m} = p, \tag{27}$$

$$\mathrm{Var}(U_{ij}) = E(U_{ij}^2) - E(U_{ij})^2 = p - p^2 = p(1-p) = \sigma^2, \tag{28}$$

and the covariance of any U_{ij} and $U_{ij'}$ pair can be derived as

$$\mathrm{Cov}(U_{ij}, U_{ij'}) = E(U_{ij}U_{ij'}) - E(U_{ij})E(U_{ij'}) = p^* - p^2 = \frac{-\sigma^2}{m-1} = \sigma^2\rho. \tag{29}$$

□

Lemma 3.2. *Let* $\mathbf{U}_{0i}$, $\mathbf{p}_0$, *and* $\mathbf{\Sigma}_0$ *be defined as in Lemma 3.1. Then it must follow that*

$$\mathbf{Y}_0 = \sqrt{n}(\bar{\mathbf{U}}_0 - \mathbf{p}_0) \xrightarrow{d} \mathcal{N}_m(\mathbf{0}, \mathbf{\Sigma}_0) \quad \text{as} \quad n \to \infty. \tag{30}$$

Proof. The expectation and variance of $\mathbf{Y}_0$ are, respectively,

$$E(\mathbf{Y}_0) = \sqrt{n}(E(\bar{\mathbf{U}}_0) - \mathbf{p}_0) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} E(\mathbf{U}_{0i}) - \mathbf{p}_0\right) = \sqrt{n}(\mathbf{p}_0 - \mathbf{p}_0) = \mathbf{0}, \tag{31}$$

$$\mathrm{Var}(\mathbf{Y}_0) = n\,\mathrm{Var}(\bar{\mathbf{U}}_0) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(\mathbf{U}_{0i}) = \mathbf{\Sigma}_0. \tag{32}$$

Therefore, given that $\mathbf{U}_{0i}$ are i.i.d. and randomly sampled, it directly follows from the multivariate Lindeberg–Lévy central limit theorem (CLT) that $\mathbf{Y}_0$ converges to an m-variate normal distribution with a mean of $\mathbf{0}$ and covariance matrix of $\mathbf{\Sigma}_0$ as n tends to infinity. □

Lemma 3.3. *Let $\mathbf{Y}_0^*$ be a vector composed of the first $m-1$ elements of $\mathbf{Y}_0$ as defined in Lemma 3.2. Then its covariance matrix $\mathbf{\Sigma}_0^*$ is invertible and*

$$(\mathbf{Y}_0^*)'(\mathbf{\Sigma}_0^*)^{-1}\mathbf{Y}_0^* \xrightarrow{d} \chi^2(m-1) \quad \text{as} \quad n \to \infty. \tag{33}$$

Proof. The covariance matrix $\mathbf{\Sigma}_0^*$ is simply an $(m-1)\times(m-1)$ submatrix of $\mathbf{\Sigma}_0$ as defined in Lemma 3.2. As such, its invertibility can be demonstrated in two ways:

- *Method 1*. The singularity of $\mathbf{\Sigma}_0$ is easily verified by observing that the sum of all elements in any kth row (or column) of $\mathbf{\Sigma}_0$ is

$$\sigma^2\left(1+\sum_{j=1}^{m-1}\rho\right) = \sigma^2[1-(m-1)/(m-1)] = 0, \tag{34}$$

  indicating linear dependence of columns (or rows). This makes sense since the above sum can also be expressed generally as

$$\sum_{j=1}^{m}\text{Cov}(U_{ik}, U_{ij}) = \text{Cov}\left(U_{ik}, \sum_{j=1}^{m} U_{ij} = L\right) = 0. \tag{35}$$

  In other words, the linear dependency in $\mathbf{\Sigma}_0$ is solely caused by the sum of all m instances of U_{ij} being determined as constant L. However, by the same token, the sum of any $m-1$ instances of U_{ij} must be random (either L or $L-1$). Therefore, by removing U_{im} from $\mathbf{Y}_0$, the sum of all elements in any kth row (or column) of $\mathbf{\Sigma}_0^*$ is

$$\sigma^2\left(1+\sum_{j=1}^{m-2}\rho\right) = \sum_{j=1}^{m-1}\text{Cov}(U_{ik}, U_{ij}) = \text{Cov}\left(U_{ik}, \sum_{j=1}^{m-1} U_{ij}\right) > 0. \tag{36}$$

  Consequently, there exist no linear combinations of columns (or rows) of $\mathbf{\Sigma}_0^*$ that sum to $\mathbf{0}$, thereby confirming that $\mathbf{\Sigma}_0^*$ is invertible.

- *Method 2*. For any $m \times m$ compound symmetric (or exchangeable) covariance matrix with the diagonal elements equal to σ^2 and the off-diagonal elements equal to δ, it can be shown (with tedious matrix algebra) that there are two possible patterns of eigenvalues λ_j:
  - if $\delta \geq 0$, then $\lambda_1 = \sigma^2 + (m-1)\delta$ and $\lambda_2 = \cdots = \lambda_m = \sigma^2 - \delta$;
  - if $\delta \leq 0$, then $\lambda_1 = \cdots = \lambda_{(m-1)} = \sigma^2 - \delta$ and $\lambda_m = \sigma^2 + (m-1)\delta$.

  (Note that if $\delta = 0$, then the matrix is diagonal and all eigenvalues are just equal to σ^2.) In the current application, $\delta = -\sigma^2/(m-1) < 0$, so the eigenvalues of $\mathbf{\Sigma}_0$ can be determined as

$$\lambda \equiv \left\{\lambda_1 = \cdots = \lambda_{(m-1)}\right\} = \sigma^2 - \frac{-\sigma^2}{m-1} = \frac{\sigma^2 m}{m-1} > 0, \tag{37}$$

$$\lambda_m = \sigma^2 + (m-1)\frac{-\sigma^2}{m-1} = 0. \tag{38}$$

In other words, the first $m-1$ eigenvalues are equal to a positive constant λ, while the last eigenvalue is equal to zero. Thus, $\boldsymbol{\Sigma}_0$ is positive semidefinite with a rank of $m-1$, meaning it cannot be inverted. On the other hand, the eigenvalues of $\boldsymbol{\Sigma}_0^*$ can be determined as

$$\lambda \equiv \left\{\lambda_1 = \cdots = \lambda_{(m-2)}\right\} = \frac{\sigma^2 m}{m-1} > 0, \tag{39}$$

$$\lambda_{(m-1)} = \sigma^2 + (m-2)\frac{-\sigma^2}{m-1} = \frac{\sigma^2}{m-1} > 0. \tag{40}$$

In other words, the first $m-2$ eigenvalues are equal to a positive constant λ, while last eigenvalue is equal to λ/m. Thus, $\boldsymbol{\Sigma}_0^*$ is positive definite with a full rank of $m-1$ and thereby invertible.

Since $\mathbf{Y}_0 \xrightarrow{d} \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Sigma}_0)$, it directly follows that $\mathbf{Y}_0^* \xrightarrow{d} \mathcal{N}_{m-1}(\mathbf{0}, \boldsymbol{\Sigma}_0^*)$. Therefore, $(\boldsymbol{\Sigma}_0^*)^{-1/2}\mathbf{Y}_0^* \xrightarrow{d} \mathcal{N}_{m-1}(\mathbf{0}, \mathbf{I})$, and by definition,

$$((\boldsymbol{\Sigma}_0^*)^{-1/2}\mathbf{Y}_0^*)'((\boldsymbol{\Sigma}_0^*)^{-1/2}\mathbf{Y}_0^*) = (\mathbf{Y}_0^*)'(\boldsymbol{\Sigma}_0^*)^{-1}\mathbf{Y}_0^* \tag{41}$$

converges to a Chi-squared distribution with $m-1$ degrees of freedom as n tends to infinity. □

Lemma 3.4. *Let $\mathbf{Y}_0$, σ^2, and λ be defined as in Lemmas 3.2–3.3. Then it must follow that*

$$\lambda^{-1}\mathbf{Y}_0'\mathbf{Y}_0 \xrightarrow{d} \chi^2(m-1) \quad \text{as} \quad n \to \infty. \tag{42}$$

Proof. Define $\boldsymbol{\Sigma}^\dagger$ to be an $m \times m$ matrix that is symmetric and idempotent with rank r. The eigendecomposition of $\boldsymbol{\Sigma}^\dagger$ is given as $\mathbf{P}\Lambda\mathbf{P}'$, where $\mathbf{P}$ is an $m \times m$ matrix of eigenvectors in the columns and $\boldsymbol{\Lambda}$ is a diagonal matrix of corresponding eigenvalues. Since $\boldsymbol{\Sigma}^\dagger$ is idempotent (i.e., $(\boldsymbol{\Sigma}^\dagger)^2 = \boldsymbol{\Sigma}^\dagger$), every eigenvalue is either 1 or 0 which sums to the number of nonzero eigenvalues, meaning $r = \text{rank}(\boldsymbol{\Sigma}^\dagger)$ and

$$\boldsymbol{\Lambda} = \begin{bmatrix} \underset{r\times r}{\mathbf{I}} & \underset{r\times(m-r)}{\mathbf{0}} \\ \underset{(m-r)\times r}{\mathbf{0}} & \underset{(m-r)\times(m-r)}{\mathbf{0}} \end{bmatrix}. \tag{43}$$

Also, since $\boldsymbol{\Sigma}^\dagger$ is symmetric (i.e., $(\boldsymbol{\Sigma}^\dagger)' = \boldsymbol{\Sigma}^\dagger$), it follows that $\mathbf{P}$ is orthogonal, meaning $\mathbf{P}' = \mathbf{P}^{-1}$ or $\mathbf{P}'\mathbf{P} = \mathbf{I}$. Next, define $\mathbf{Z}$ to be a random m-vector that converges in distribution to $\mathcal{N}_m(\mathbf{0}, \mathbf{I})$. Due to the aforementioned properties of $\boldsymbol{\Sigma}^\dagger$, multiplying it to $\mathbf{Z}$ projects the m-vector onto an orthogonal r-dimensional subspace, resulting in $\boldsymbol{\Sigma}^\dagger\mathbf{Z} \xrightarrow{d} \mathcal{N}_r(\mathbf{0}, \boldsymbol{\Sigma}^\dagger)$. Furthermore, note that $\mathbf{P}'\mathbf{Z} \xrightarrow{d} \mathcal{N}_m(\mathbf{P}'\mathbf{0}, \mathbf{P}'\mathbf{I}\mathbf{P}) = \mathcal{N}_m(\mathbf{0}, \mathbf{I})$. In other words, $\mathbf{P}'\mathbf{Z}$ is composed of independent Z_j that, $\forall\ j \in \{1, \ldots, m\}$, converge to standard normality. Therefore, the squared length of the projection is

$$||\boldsymbol{\Sigma}^\dagger\mathbf{Z}||^2 = (\boldsymbol{\Sigma}^\dagger\mathbf{Z})'(\boldsymbol{\Sigma}^\dagger\mathbf{Z}) = \mathbf{Z}'\boldsymbol{\Sigma}^\dagger\mathbf{Z} = (\mathbf{P}'\mathbf{Z})'\boldsymbol{\Lambda}(\mathbf{P}'\mathbf{Z}) = \sum_{j=1}^{r} Z_j^2, \tag{44}$$

which by definition converges to a χ^2 distribution with r degrees of freedom.

To take advantage of the results above, $\mathbf{Y}_0$ needs to be transformed in such a way that $\boldsymbol{\Sigma}_0$ is correspondingly transformed into an orthogonal projection matrix with the same rank of $r = m - 1$:

$$\mathbf{A}\mathbf{Y}_0 = \boldsymbol{\Sigma}^{\dagger}\mathbf{Z} \xrightarrow{d} \mathcal{N}_{(m-1)}(\mathbf{0}, \boldsymbol{\Sigma}^{\dagger} = \mathbf{A}\boldsymbol{\Sigma}_0\mathbf{A}'). \tag{45}$$

To find the appropriate transformation matrix $\mathbf{A}$, the trick is to make the trace of $\boldsymbol{\Sigma}^{\dagger}$ equal to the rank of $\boldsymbol{\Sigma}_0$ since the trace and rank of an idempotent matrix are equivalent. The diagonal entries of $\boldsymbol{\Sigma}_0$ are all equal, so this can be achieved by setting the rank of $\boldsymbol{\Sigma}_0$ equal to the trace of $\boldsymbol{\Sigma}_0$ multiplied by some constant t, or $m - 1 = t(m\sigma^2)$. Then solving for t:

$$t = \frac{m-1}{m\sigma^2} = \frac{1}{\lambda}, \tag{46}$$

which happens to be the inverse of the nonzero eigenvalue of $\boldsymbol{\Sigma}_0$. Hence, $\mathbf{A}$ is simply the scalar matrix $\sqrt{t}\mathbf{I}$, which can be verified as follows:

$$\mathbf{A}\mathbf{Y}_0 = (\sqrt{t}\mathbf{I})\mathbf{Y}_0 = \sqrt{t}\mathbf{Y}_0, \tag{47}$$

$$\boldsymbol{\Sigma}^{\dagger} = \mathbf{A}\boldsymbol{\Sigma}_0\mathbf{A}' = (\sqrt{t}\mathbf{I})\boldsymbol{\Sigma}_0(\sqrt{t}\mathbf{I})' = t\boldsymbol{\Sigma}_0 = \mathbf{I} - \mathbf{q}\mathbf{q}', \tag{48}$$

where $\mathbf{q} = [1/\sqrt{m}, \ldots, 1/\sqrt{m}]'$. Note that $\boldsymbol{\Sigma}^{\dagger}$ is symmetric with diagonal and off-diagonal entries equal to $(m-1)/m$ and $-1/m$, respectively. Furthermore, its idempotence can be checked as follows:

$$(\boldsymbol{\Sigma}^{\dagger})^2 = (\mathbf{I} - \mathbf{q}\mathbf{q}')(\mathbf{I} - \mathbf{q}\mathbf{q}') = \mathbf{I}^2 - 2\mathbf{I}\mathbf{q}\mathbf{q}' + \mathbf{q}\mathbf{q}'\mathbf{q}\mathbf{q}' = \mathbf{I} - \mathbf{q}\mathbf{q}' = \boldsymbol{\Sigma}^{\dagger}, \tag{49}$$

whose rank is equal to its trace of $m - 1$. Therefore,

$$||\mathbf{A}\mathbf{Y}_0||^2 = (\mathbf{A}\mathbf{Y}_0)'\mathbf{A}\mathbf{Y}_0 = t\mathbf{Y}_0'\mathbf{Y}_0 = \lambda^{-1}\mathbf{Y}_0'\mathbf{Y}_0, \tag{50}$$

which converges to a χ^2 distribution with $m - 1$ degrees of freedom as n tends to infinity. □

Lemma 3.5. *Let* $\mathbf{Y}_0$, $\mathbf{Y}_0^*$, $\boldsymbol{\Sigma}_0^*$, *and* λ *be defined as in Lemmas 3.2–3.4. Then it must follow that*

$$(\mathbf{Y}_0^*)'(\boldsymbol{\Sigma}_0^*)^{-1}\mathbf{Y}_0^* = \lambda^{-1}\mathbf{Y}_0'\mathbf{Y}_0. \tag{51}$$

Proof. The eigendecomposition of $\boldsymbol{\Sigma}_0^*$ is given as $\mathbf{P}^*\Lambda^*(\mathbf{P}^*)'$, where $\mathbf{P}^*$ is an orthogonal matrix of eigenvectors in the columns and $\boldsymbol{\Lambda}^*$ is a diagonal matrix of corresponding eigenvalues as follows:

$$\mathbf{P}^* = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1,(m-2)} & x_{1,(m-1)} \\ x_{21} & x_{22} & \cdots & x_{2,(m-2)} & x_{2,(m-1)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{(m-2),1} & x_{(m-2),2} & \cdots & x_{(m-2),(m-2)} & x_{(m-2),(m-1)} \\ x_{(m-1),1} & x_{(m-1),2} & \cdots & x_{(m-1),(m-2)} & x_{(m-1),(m-1)} \end{bmatrix}, \quad \boldsymbol{\Lambda}^* = \lambda \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1/m \end{bmatrix},$$

where $\lambda = \sigma^2 m/(m-1)$ as derived in the proof of Lemma 3.3. Given that the columns and rows of $\mathbf{P}^*$ are orthonormal, it follows that

$$\sum_{i=1}^{m-1} x_{ij}^2 = \sum_{j=1}^{m-1} x_{ij}^2 = 1, \quad \sum_{i=1}^{m-1} x_{ij} x_{ij'} = \sum_{j=1}^{m-1} x_{ij} x_{i'j} = 0.$$

Furthermore, by solving the eigenvalue equation $\mathbf{\Sigma}_0^* \mathbf{x}_{(m-1)} = (\lambda/m)\mathbf{x}_{(m-1)}$ for the eigenvector $\mathbf{x}_{(m-1)}$ (last column of $\mathbf{P}^*$), it is easily verified that

$$x_{1,(m-1)} = x_{2,(m-1)} = \cdots = x_{(m-2),(m-1)} = x_{(m-1),(m-1)}. \tag{52}$$

Hence, by utilizing the above identities, the following quantities are readily derived:

$$x_{i,(m-1)}^2 = \frac{1}{m-1}, \quad \sum_{j=1}^{m-2} x_{ij}^2 = \frac{m-2}{m-1}, \quad \sum_{j=1}^{m-2} x_{ij} x_{i'j} = \frac{-1}{m-1}.$$

The inverse of $\mathbf{\Sigma}_0^*$ is

$$(\mathbf{\Sigma}_0^*)^{-1} = (\mathbf{P}^* \Lambda^* (\mathbf{P}^*)')^{-1} = \mathbf{P}^* (\mathbf{\Lambda}^*)^{-1} (\mathbf{P}^*)', \tag{53}$$

where

$$(\mathbf{\Lambda}^*)^{-1} = \lambda^{-1} \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & m \end{bmatrix}. \tag{54}$$

Therefore, the diagonal and off-diagonal entries of $(\mathbf{\Sigma}_0^*)^{-1}$ are obtained as, respectively,

$$\lambda^{-1} \left(\sum_{j=1}^{m-2} x_{ij}^2 + m x_{i,(m-1)}^2 \right) = \lambda^{-1} \left(\frac{m-2}{m-1} + \frac{m}{m-1} \right) = \frac{2}{\lambda}, \tag{55}$$

$$\lambda^{-1} \left(\sum_{j=1}^{m-2} x_{ij} x_{i'j} + m x_{i,(m-1)}^2 \right) = \lambda^{-1} \left(\frac{-1}{m-1} + \frac{m}{m-1} \right) = \frac{1}{\lambda}, \tag{56}$$

which can be expressed as

$$(\mathbf{\Sigma}_0^*)^{-1} = \lambda^{-1}(\mathbf{I} + \mathbf{J}), \tag{57}$$

where $\mathbf{J}$ is a matrix of ones. Finally, the equivalence is established as follows:

$$(\mathbf{Y}_0^*)'(\mathbf{\Sigma}_0^*)^{-1}\mathbf{Y}_0^* = \lambda^{-1}(\mathbf{Y}_0^*)'(\mathbf{I} + \mathbf{J})\mathbf{Y}_0^*$$

$$\begin{aligned}
&= \lambda^{-1}[(\mathbf{Y}_0^*)'\mathbf{Y}_0^* + (\mathbf{Y}_0^*)'\mathbf{J}\mathbf{Y}_0^*] \\
&= \lambda^{-1}\left[n\sum_{j=1}^{m-1}(\bar{U}_{0j}-p)^2 + n\left(\sum_{j=1}^{m-1}(\bar{U}_{0j}-p)\right)^2\right] \\
&= \lambda^{-1}\left[n\sum_{j=1}^{m-1}(\bar{U}_{0j}-p)^2 + n(p-\bar{U}_{0m})^2\right] \\
&= \lambda^{-1} n\sum_{j=1}^{m}(\bar{U}_{0j}-p)^2 \\
&= \lambda^{-1}\mathbf{Y}_0'\mathbf{Y}_0.
\end{aligned}$$

□

Theorem 3.6. *Suppose a random sample of n examinees are each* ***randomly*** *administered L items from a pool of m items, resulting in a total of $N = nL$ item exposures. Let*

$$\psi = (m-1)/(m-L), \quad \beta = \psi m(n-1), \quad \gamma = \psi(m-N).$$

Furthermore, let $\bar{R}_0$ be the average test overlap rate as defined in Lemma 2.1. Then it must follow that

$$Q_0 = \beta\bar{R}_0 + \gamma \xrightarrow{d} \chi^2(m-1) \quad \text{as} \quad n \to \infty. \tag{58}$$

Proof. According to Lemma 3.1, the variance of U_{ij} is

$$\sigma^2 = p(1-p) = \frac{L}{m}\left(1-\frac{L}{m}\right), \tag{59}$$

so it follows from the proofs of Lemmas 3.3–3.4 that the inverse of the nonzero eigenvalue of $\mathbf{\Sigma}_0$ is

$$\lambda^{-1} = \frac{m-1}{m\sigma^2} = \frac{m-1}{m(L/m)(1-L/m)} = \frac{m(m-1)}{L(m-L)} = \frac{\psi}{L/m}. \tag{60}$$

Then using Lemma 3.5, define the Q_0 statistic as

$$Q_0 = (\mathbf{Y}_0^*)'(\mathbf{\Sigma}_0^*)^{-1}\mathbf{Y}_0^* = \lambda^{-1}\mathbf{Y}_0'\mathbf{Y}_0 = \frac{\psi n\sum_{j=1}^{m}(\bar{U}_{0j}-L/m)^2}{L/m}, \tag{61}$$

where the sum of squares quantity can be expanded as

$$\sum_{j=1}^{m}(\bar{U}_{0j}-L/m)^2 = \sum_{j=1}^{m}\left(\bar{U}_{0j}^2 - 2\bar{U}_{0j}L/m + L^2/m^2\right) = \sum_{j=1}^{m}\bar{U}_{0j}^2 - L^2/m. \tag{62}$$

Note that when $L = 1$, Q_0 reduces to the familiar Pearson's χ^2 test statistic. Lastly, Lemma 2.1 enables the derivation of $\sum_{j=1}^{m}\bar{U}_{0j}^2$ as a function of $\bar{R}_0$ as follows:

$$\sum_{j=1}^{m} \bar{U}_{0j}^2 = \frac{L}{n}\left[\bar{R}_0(n-1)+1\right]. \tag{63}$$

Therefore, Q_0 can be equivalently expressed as the following linear transformation of $\bar{R}_0$:

$$Q_0 = \frac{\psi n\left(\sum_{j=1}^{m} \bar{U}_{0j}^2 - L^2/m\right)}{L/m} = \psi m(n-1)\bar{R}_0 + \psi(m-N) \tag{64}$$

which, by Lemma 3.3 or Lemma 3.4, converges to a Chi-squared distribution with $m-1$ degrees of freedom as n tends to infinity. □

Corollary 3.6.1. *Suppose a random sample of n examinees are each* ***randomly*** *administered L items from a pool of m items. Let ψ and $\bar{R}_0$ be defined as in Theorem 3.6, and let*

$$\phi = m(n-1)/n, \quad \tau = m/n - L.$$

Then it must follow that

$$G = \phi\bar{R}_0 + \tau \xrightarrow{d} \Gamma\left(\frac{m-1}{2}, \frac{2}{\psi n}\right) \quad \text{as} \quad n \to \infty. \tag{65}$$

Proof. G is just Q_0 (as defined in Theorem 3.6) scaled by $(\psi n)^{-1}$ as follows:

$$G = \phi\bar{R}_0 + \tau = Q_0/(\psi n) = \frac{\sum_{j=1}^{m}(\bar{U}_{0j} - L/m)^2}{L/m}. \tag{66}$$

G is the "χ^2" statistic that was originally proposed by Chang and Ying (1999) as a measure of item pool utilization or, more specifically, the imbalance of exposures across items in the pool. A larger value of G indicates greater imbalance of item exposure rates, and its computation as a function of $\bar{R}_0$ was first reported by Choe, Kern, and Chang (2018). However, G does not actually converge to a χ^2 distribution. Instead, it is well known that multiplying a constant κ to a Chi-squared random variable with ν degrees of freedom results in a gamma random variable with a shape parameter of $\nu/2$ and a scale parameter of 2κ. Therefore, given that $\kappa = (\psi n)^{-1}$ and $\nu = m-1$, G converges to a gamma distribution with shape and scale parameters of $(m-1)/2$ and $2/(\psi n)$, respectively, as n tends to infinity. □

4. The Focal Distribution of Average Test Overlap Rate

Next, consider the assumption that a fixed number of items are selected adaptively for every examinee. By design, adaptive testing selects items in a non-random fashion according to an individual's interim performance based on θ, usually resulting in an imbalanced usage of items across the pool even with a provision for exposure control.

Lemma 4.1. *Suppose a random sample of n examinees with same ability θ are each* ***adaptively*** *administered L items from a pool of m items. Then, $\forall\ i \in \{1, \ldots, n\}$, $\mathbf{U}_i|\theta$ are i.i.d. as m-dimensional Bernoulli with expectation $\mathbf{p}_1$ and covariance $\boldsymbol{\Sigma}_1$ as follows:*

$$\mathbf{p}_1 = [p_1 \ \ p_2 \ \ \cdots \ \ p_m]', \quad \boldsymbol{\Sigma}_1 = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1m} & \sigma_{2m} & \cdots & \sigma_m^2 \end{bmatrix},$$

where $0 \leq \sigma_j^2 \leq 0.25$ *and* $-0.25 \leq \sigma_{jj'} \leq 0$.

Proof. Given a fixed ability level, an ith trial is a non-uniform but still random sampling without replacement of L items from a pool of m items. In other words, item exposures are i.i.d. strictly conditional on θ. As in Method 1 in the proof of Lemma 3.1, define Ω as the set of every possible ordered sequence of L item exposures for a given trial, with $\omega_s \in \Omega$ representing a unique elementary event $s \in \{1, \ldots, |\Omega|\}$. Additionally, define $\mathbf{q} = [q_1 \ q_2 \ \cdots \ q_m]'$ to be a vector of initial selection probabilities for any given trial. Then, without loss of generality, the probability of a specific instance ω_3 given θ is computed in the following recursive fashion:

$$P(\omega_3|\theta) = q_4 \cdot \frac{q_7}{1 - q_4} \cdot \frac{q_2}{1 - q_4 - q_7} \cdots \tag{67}$$

until the Lth selection. Note that for each subsequent selection, the initial probability is divided by one minus the probabilities of previous selections for normalization. Then, defining $\mathcal{C}_j \subseteq \Omega$ to be the subset of all instances in which a particular item j is selected, exposure probability of item j is generally given as

$$p_j := P(U_{ij} = 1|\theta) = \sum_{w_s \in \mathcal{C}_j} P(w_s|\theta). \tag{68}$$

Likewise, defining $\mathcal{D}_{jj'} \subseteq \Omega$ to be the subset of all instances in which a pair of distinct items j and j' are selected, the joint exposure probability of items j and j' is generally given as

$$p_{jj'} := P(U_{ij} = 1, U_{ij'} = 1|\theta) = \sum_{w_s \in \mathcal{D}_{jj'}} P(w_s|\theta). \tag{69}$$

Thus, the expected value and variance of U_{ij} given θ are, respectively,

$$E(U_{ij}|\theta) = P(U_{ij} = 1|\theta) = p_j, \tag{70}$$

$$\mathrm{Var}(U_{ij}|\theta) = E(U_{ij}^2|\theta) - E(U_{ij}|\theta)^2 = p_j - p_j^2 = p_j(1 - p_j) = \sigma_j^2, \tag{71}$$

and the covariance of U_{ij} and $U_{ij'}$ given θ is

$$\mathrm{Cov}(U_{ij}, U_{ij'}|\theta) = E(U_{ij}U_{ij'}|\theta) - E(U_{ij}|\theta)E(U_{ij'}|\theta) = p_{jj'} - p_j^2 = \sigma_{jj'}. \tag{72}$$

It is trivial to show that $\min \sigma_j^2 = 0$ when $p_j = 0$ or 1 and $\max \sigma_j^2 = 0.25$ when $p_j = 0.5$. Furthermore, the proof of Lemma 3.3 revealed that $\sigma_{j'}^2 + \sum_{j \neq j'} \sigma_{jj'} = 0$, and it can also be shown that $\sigma_{jj'} \leq 0$. Therefore, it must hold that $\min \sigma_{jj'} = -0.25$. □

Lemma 4.2. *Suppose a random sample of n examinees with individual abilities θ_i is drawn from the sample space Θ (i.e., $\theta_i \in \Theta$), and they are each **adaptively** administered L items from a pool of m items. For any ith examinee, let $\mathbf{U}_i|\theta_i$, $\mathbf{p}_{1i}$, and $\mathbf{\Sigma}_{1i}$ be individually defined as in Lemma 4.1. The unconditional expectation and variance of $\mathbf{U}_i$ can then be expressed as, respectively,*

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{p}_{1i} \to \mathbf{p} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}\mathbf{\Sigma}_{1i} + \mathbf{\Sigma}_p \to \mathbf{\Sigma} \quad \text{as} \quad n \to \infty, \tag{73}$$

where $\mathbf{\Sigma}_p$ is the covariance matrix of $\mathbf{p}_{1i}$ (i.e., $\mathrm{Var}(\mathbf{p}_{1i}) = \mathbf{\Sigma}_p$).

Proof. The law of total expectation states that

$$\mathbf{p} := E(\mathbf{U}_i) = E[E(\mathbf{U}_i|\theta_i)] = E(\mathbf{p}_{1i}), \tag{74}$$

and the law of total variance states that

$$\mathbf{\Sigma} := \mathrm{Var}(\mathbf{U}_i) = E[\mathrm{Var}(\mathbf{U}_i|\theta_i)] + \mathrm{Var}[E(\mathbf{U}_i|\theta_i)] = E(\mathbf{\Sigma}_{1i}) + \mathrm{Var}(\mathbf{p}_{1i}). \tag{75}$$

Essentially, $\mathbf{p}$ and $\mathbf{\Sigma}$ can be thought of as the expectation and variance of $\mathbf{U}_i$ with θ_i averaged out. Note that $\mathbf{p}_{1i}$ and $\mathbf{\Sigma}_{1i}$ are random quantities that are solely dependent on θ_i randomly sampled from Θ. In other words, both $\mathbf{p}_{1i}$ and $\mathbf{\Sigma}_{1i}$ are, respectively, i.i.d. $\forall\ i \in \{1, \ldots, n\}$. Therefore, the law of large numbers (LLN) guarantees the converge of their sample means to $E(\mathbf{p}_{1i})$ and $E(\mathbf{\Sigma}_{1i})$, respectively. Moreover, the existence of a finite $\mathbf{p}$ can be inferred by noting that, as a vector of probabilities, the elements of $\mathbf{p}_{1i}$ are restricted to a range from 0 to 1. Likewise, it can be inferred from Lemma 4.1 that a finite $\mathbf{\Sigma}$ exists, because the absolute value of the elements of $\mathbf{\Sigma}_{1i}$ can only range from 0 to 0.25. □

Lemma 4.3. *Let $\mathbf{U}_i^*$ be the first $m-1$ elements of $\mathbf{U}_i$, as defined in Lemma 4.2, with expectation $\mathbf{p}^*$ and covariance $\mathbf{\Sigma}^*$. Furthermore, let $\mathbf{p}_0^*$ be the first $m-1$ elements of $\mathbf{p}_0$ as defined in Lemma 3.1. Then, given $\boldsymbol{\mu}^* = \sqrt{n}(\mathbf{p}^* - \mathbf{p}_0^*)$, it must follow that*

$$\mathbf{Y}^* = \sqrt{n}(\bar{\mathbf{U}}^* - \mathbf{p}_0^*) \xrightarrow{d} \mathcal{N}_m(\boldsymbol{\mu}^*, \mathbf{\Sigma}^*) \quad \text{as} \quad n \to \infty. \tag{76}$$

Proof. Having removed the dependency on θ_i as explained in the proof of Lemma 4.2, the unconditional $\mathbf{U}_i^*$ can be perceived as i.i.d. random variables drawn from a distribution with an expectation of $\mathbf{p}^*$ and a covariance matrix of $\mathbf{\Sigma}^*$. The expectation and variance of $\mathbf{Y}^*$ are then, respectively,

$$E(\mathbf{Y}^*) = \sqrt{n}(E(\bar{\mathbf{U}}^*) - \mathbf{p}_0^*) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}E(\mathbf{U}_i^*) - \mathbf{p}_0^*\right) = \sqrt{n}\left(\mathbf{p}^* - \mathbf{p}_0^*\right) = \boldsymbol{\mu}^*, \tag{77}$$

$$\mathrm{Var}(\mathbf{Y}^*) = n\,\mathrm{Var}(\bar{\mathbf{U}}^*) = n\left(\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(\mathbf{U}_i^*)\right) = \mathbf{\Sigma}^*. \tag{78}$$

Therefore, the multivariate Lindeberg–Lévy CLT implies that $\mathbf{Y}^*$ must converge to an m-variate normal distribution with a mean of $\boldsymbol{\mu}^*$ and covariance matrix of $\mathbf{\Sigma}^*$ as n tends to infinity. □

Theorem 4.4. *Suppose a random sample of n examinees are each **adaptively** administered L items from a pool of m items. Let β and γ be defined as in Theorem 3.6. Furthermore, let*

$$M = \mathrm{tr}((\boldsymbol{\Sigma}_0^*)^{-1}\boldsymbol{\Sigma}^*) + (\boldsymbol{\mu}^*)'(\boldsymbol{\Sigma}_0^*)^{-1}\boldsymbol{\mu}^*,$$
$$V = 2\,\mathrm{tr}([(\boldsymbol{\Sigma}_0^*)^{-1}\boldsymbol{\Sigma}^*]^2) + 4\,(\boldsymbol{\mu}^*)'(\boldsymbol{\Sigma}_0^*)^{-1}\boldsymbol{\Sigma}^*(\boldsymbol{\Sigma}_0^*)^{-1}\boldsymbol{\mu}^*,$$

where $\boldsymbol{\Sigma}_0^*$ is defined as in Lemma 3.3, whereas $\boldsymbol{\Sigma}^*$ and $\boldsymbol{\mu}^*$ are defined as in Lemma 4.3. Then, given the average test overlap rate $\bar{R}$ as defined in Lemma 2.1, it must follow that

$$Q = \beta\bar{R} + \gamma \xrightarrow{d} \mathcal{N}(M, V) \quad \text{as} \quad n \to \infty \quad \text{and} \quad m \to \infty. \tag{79}$$

Proof. Consider the following statistic,

$$Q = (\mathbf{Y}^*)'(\boldsymbol{\Sigma}_0^*)^{-1}\mathbf{Y}^*, \tag{80}$$

where $\mathbf{Y}^*$ is defined as in Lemma 4.3. Note that Q is algebraically equivalent to baseline Q_0, so the proof of Theorem 3.6 directly applies here to derive Q as a function of $\bar{R}$ as stated. However, the distribution of Q is generally not the same as that of Q_0. In fact, the precise limiting distribution of Q is unknown with respect to n alone. Nevertheless, since Q is a quadratic form, its expectation can be expressed as

$$M := E(Q) = \mathrm{tr}((\boldsymbol{\Sigma}_0^*)^{-1}\boldsymbol{\Sigma}^*) + (\boldsymbol{\mu}^*)'(\boldsymbol{\Sigma}_0^*)^{-1}\boldsymbol{\mu}^*. \tag{81}$$

Furthermore, since $\mathbf{Y}^*$ is asymptotically normal, the variance of Q can be approximated as

$$V := \mathrm{Var}(Q) = 2\,\mathrm{tr}([(\boldsymbol{\Sigma}_0^*)^{-1}\boldsymbol{\Sigma}^*]^2) + 4\,(\boldsymbol{\mu}^*)'(\boldsymbol{\Sigma}_0^*)^{-1}\boldsymbol{\Sigma}^*(\boldsymbol{\Sigma}_0^*)^{-1}\boldsymbol{\mu}^*. \tag{82}$$

(Rencher & Schaalje, 2008). Lastly, Q can be interpreted as a sum of $m-1$ independent random variables. Therefore, CLT implies that Q converges to normality with mean M and variance V as both n and m tend to infinity. □

Example 4.4.1. (*Confidence Interval*) Define ξ to be the expected mean of test overlap rate, i.e., $\xi = E(\bar{R})$. Then an approximate confidence interval for ξ can be constructed by obtaining one for M and transforming it to the scale of ξ as follows:

$$\frac{Q - z_{\alpha/2}\sqrt{V} - \gamma}{\beta} \le \xi \le \frac{Q + z_{\alpha/2}\sqrt{V} - \gamma}{\beta}, \tag{83}$$

where $z_{\alpha/2}$ is a standard normal quantile corresponding to the desired level of Type I error (α). Note that the calculation of the standard error $\sqrt{V}$ requires both $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$, which are unknown in practice. Instead, provided a reasonably large sample that is representative of Θ, their sample estimates can be obtained as

$$\hat{\boldsymbol{\mu}}^* = \mathbf{Y}^*, \quad \hat{\boldsymbol{\Sigma}}^* = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1,m-1} \\ s_{12} & s_2^2 & \cdots & s_{2,m-1} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1,m-1} & s_{2,m-1} & \cdots & s_{m-1}^2 \end{bmatrix},$$

where s_j^2 is the sample variance of U_{ij} and $s_{jj'}$ is the sample covariance between U_{ij} and $U_{ij'}$. The variance term can then be estimated as

$$\hat{V} = 2\,\mathrm{tr}([(\boldsymbol{\Sigma}_0^*)^{-1}\hat{\boldsymbol{\Sigma}}^*]^2) + 4\,(\hat{\boldsymbol{\mu}}^*)'(\boldsymbol{\Sigma}_0^*)^{-1}\hat{\boldsymbol{\Sigma}}^*(\boldsymbol{\Sigma}_0^*)^{-1}\hat{\boldsymbol{\mu}}^*. \tag{84}$$

TABLE 2.
Key quantities and corresponding descriptions or algebraic expressions

Quantity	Description/expression
n	Sample size (# of examinees)
m	Pool size (# of items)
L	Test length
U_{ij}	1 for exposure of item j to examinee i, 0 otherwise
$\bar{U}_j$	$\sum_{i=1}^{n} U_{ij}/n$
s_j^2	$\sum_{i=1}^{n}(U_{ij}-\bar{U}_j)^2/(n-1)$
$s_{jj'}$	$\sum_{i=1}^{n}(U_{ij}-\bar{U}_j)(U_{ij'}-\bar{U}_{j'})/(n-1)$
$\bar{R}$	$n/(n-1)\sum_{j=1}^{m}\bar{U}_j^2/L-(n-1)^{-1}$
p	L/m
σ^2	$p(1-p)$
λ	$\sigma^2 m/(m-1)$
ψ	$(m-1)/(m-L)$
β	$\psi m(n-1)$
γ	$\psi(m-nL)$
$\mathbf{I}$	$(m-1)\times(m-1)$ identity matrix
$\mathbf{J}$	$(m-1)\times(m-1)$ matrix of ones
$\bar{\mathbf{U}}^*$	$[\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_{m-1}]$
$\mathbf{p}_0^*$	$[p, p, \ldots, p]'$
$\hat{\boldsymbol{\mu}}^*$	$\sqrt{n}(\bar{\mathbf{U}}^*-\mathbf{p}_0^*)$
$(\boldsymbol{\Sigma}_0^*)^{-1}$	$\lambda^{-1}(\mathbf{I}+\mathbf{J})$
$\hat{\boldsymbol{\Sigma}}^*$	$\begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1,m-1} \\ s_{12} & s_2^2 & \cdots & s_{2,m-1} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1,m-1} & s_{2,m-1} & \cdots & s_{m-1}^2 \end{bmatrix}$
Q	$\beta\bar{R}+\gamma$
$\hat{V}$	$2\,\mathrm{tr}([(\boldsymbol{\Sigma}_0^*)^{-1}\hat{\boldsymbol{\Sigma}}^*]^2)+4\,(\hat{\boldsymbol{\mu}}^*)'(\boldsymbol{\Sigma}_0^*)^{-1}\hat{\boldsymbol{\Sigma}}^*(\boldsymbol{\Sigma}_0^*)^{-1}\hat{\boldsymbol{\mu}}^*$
$\hat{Z}$	$(Q^{(1)}-Q^{(2)})(\hat{V}^{(1)}+\hat{V}^{(2)})^{-1/2}$

Example 4.4.2. (*Hypothesis Testing*) The average test overlap rates between two independent CAT cycles can be compared using the following asymptotically normal test statistic:

$$Z=\frac{Q^{(1)}-Q^{(2)}}{\sqrt{V^{(1)}+V^{(2)}}}=\frac{(\beta^{(1)}\bar{R}^{(1)}+\gamma^{(1)})-(\beta^{(2)}\bar{R}^{(2)}+\gamma^{(2)})}{\sqrt{V^{(1)}+V^{(2)}}}. \tag{85}$$

If the corresponding values of L, m, and n are equal for both samples (i.e., $\beta^{(1)} = \beta^{(2)} = \beta$ and $\gamma^{(1)} = \gamma^{(2)}$), then the test statistic can be simplified to

$$Z = \frac{\beta(\bar{R}^{(1)} - \bar{R}^{(2)})}{\sqrt{V^{(1)} + V^{(2)}}}. \tag{86}$$

Like the previous example, the standard error can be estimated as $\sqrt{\hat{V}^{(1)} + \hat{V}^{(2)}}$ to compute an estimate of Z, or $\hat{Z}$. As a convenient reference for computations, a list of key quantities and their corresponding descriptions or algebraic expressions is given in Table 2.

Under the null case of $Q^{(1)}$ and $Q^{(2)}$ having identical distributions, $\hat{Z}$ should converge to standard normality as n and m increase. To empirically verify this, CAT simulations were run with a basic MFI algorithm using various sample sizes ($n = 20, 60, 100$) and pool sizes ($m = 25, 50$) with test length fixed at $L = 10$. For each of the six $\{n, m\}$ conditions, 1000 $\hat{Z}$ values were simulated and plotted as empirical density curves in Fig. 1. Observe that the distribution of $\hat{Z}$ is nearly standard normal for n and m as small as 100 and 50, respectively.

5. CAT Simulation Comparing Two Exposure Control Methods

A practical application of Example 4.4.2 is comparing the performances of two exposure control methods in CAT. As an illustration, a realistic CAT simulation study was designed with the following specifications:

- Examinees: 500 simulees ($n = 500$) whose θ_i values were randomly sampled with replacement from a pool of recent candidates in an operational testing program;
- Item pool: 250 items ($m = 250$) spanning several content domains, expressly assembled for simulation purposes from a much larger operational bank calibrated according to 3PLM;
- Test blueprint: mini-test fixed at 10 items ($L = 10$), with at least one item from each content domain;
- Response: probabilistic generation of dichotomous answer (either correct or incorrect) for each simulee i on each item j based on $P_j(\theta_i)$;
- Item selection criterion: maximum Fisher information (MFI);
- Ability estimation: MLE with fences at $[-5, 5]$ (MLEF; Han, 2016); initial $\hat{\theta}_i$ randomly generated from Uniform $[-0.5, 0.5]$ for selecting first item.

Strictly under these conditions, two distinct exposure control methods were implemented on separate runs and compared. The first method was the classic Sympson–Hetter technique (Sympson & Hetter, 1985), in which item exposure is limited stochastically by setting a probability on each item's actual administration ($\mathcal{A}$) given its selection ($\mathcal{B}$) by the CAT algorithm. This probability is predetermined for an item j as $P(\mathcal{A}_j|\mathcal{B}_j) = r/P(\mathcal{B}_j)$, where r is a desired maximum exposure rate for every item and $P(\mathcal{B}_j)$ is a converged value obtained through multiple iterations of CAT simulations. If a selected item is decided not to be administered, then the algorithm repeats the process by selecting another item with the next highest information. The second method was a variation of the so-called randomesque technique (Kingsbury & Zara 1989), in which one item is randomly administered from a set of eligible items with the highest information at the current interim $\hat{\theta}_i$. The size of the eligible set was determined by taking a specified percentage of available items within a chosen content domain, which was graduated based on item position (e.g., from top 50% for the first item down to top 1% for the last item).

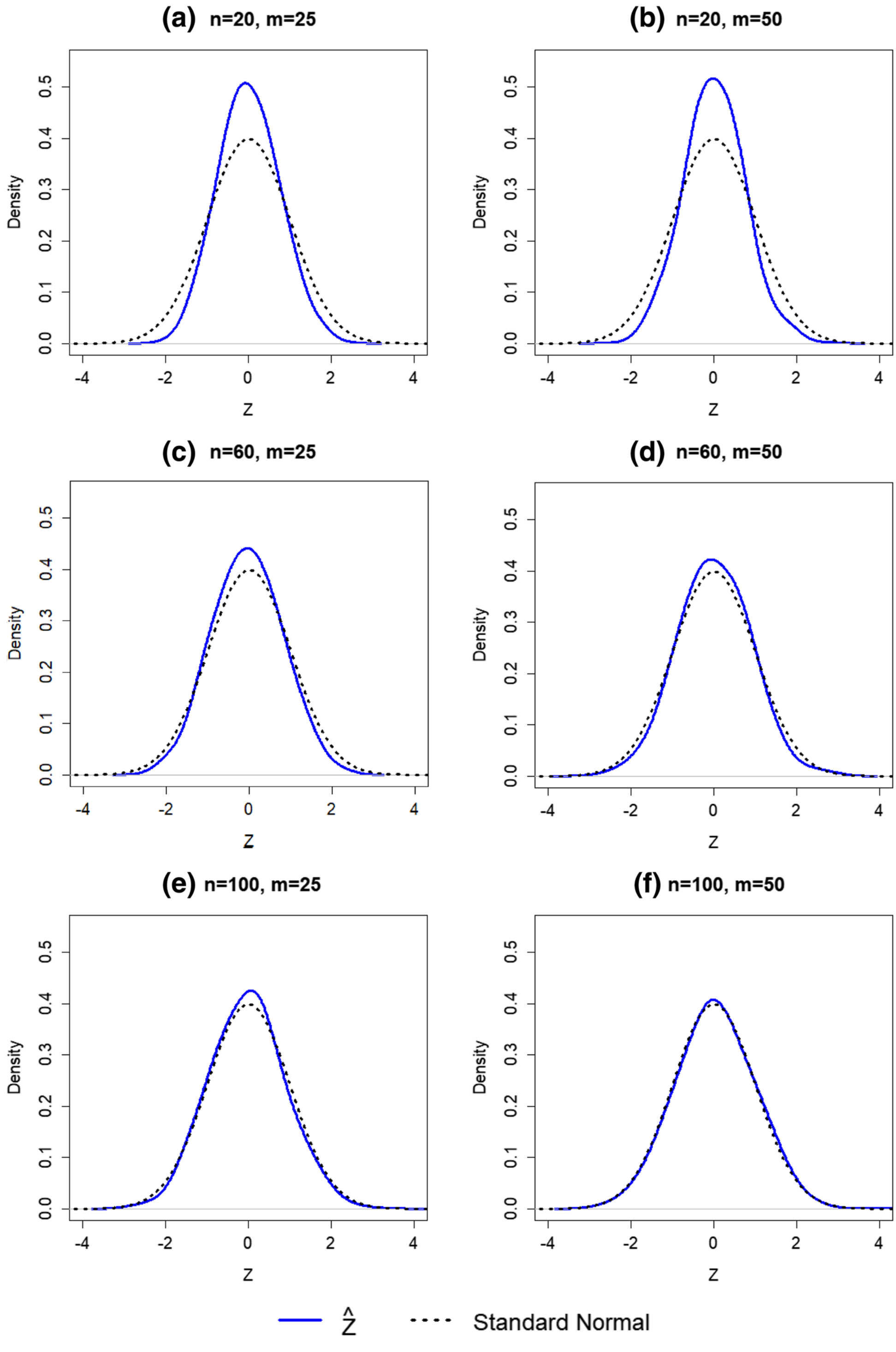

FIGURE 1.
Simulated density plots of $\hat{Z}$ under the null (solid curve) are shown according to $n = 20, 60, 100$ and $m = 25, 50$ with $L = 10$. For given n and m, 1000 pairs of CAT replications were run with a basic MFI algorithm. Larger n and m result in greater convergence to standard normality (dotted curve).

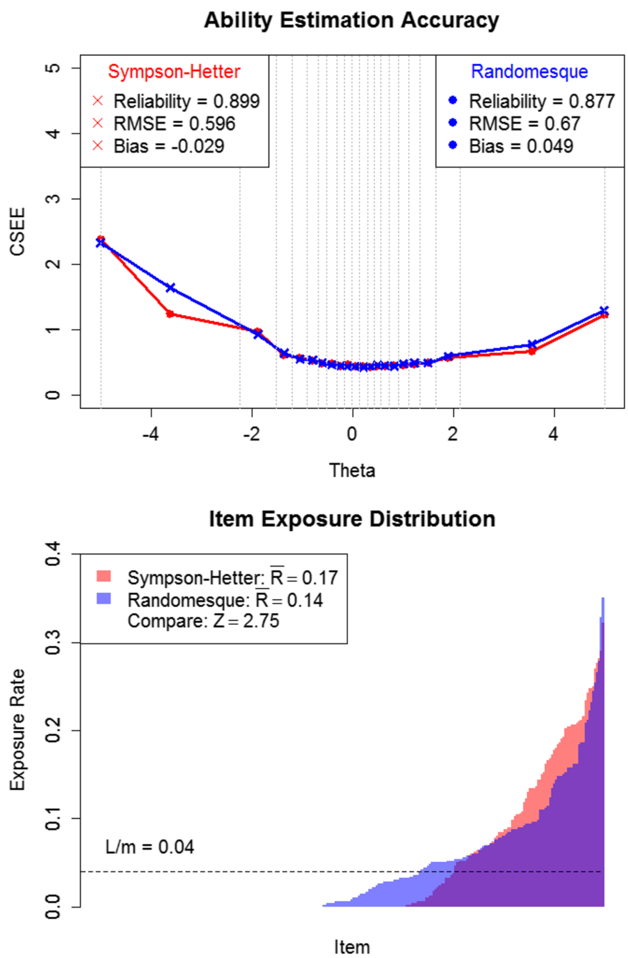

FIGURE 2.
These plots display the results of a CAT simulation study comparing the Sympson–Hetter and randomesque methods of exposure control for a given item pool and test specifications. Sympson–Hetter seems to afford slightly better ability estimations, specifically at the extremes (top plot). However, randomesque results in about 0.03 lower average test overlap rate that is statistically significant at $\alpha = 0.01$ (bottom plot).

The comparative performance results are presented in Fig. 2. The top plot shows the conditional standard error of estimation (CSEE) across the θ range, which was computed as

$$\mathrm{CSEE}(\theta_p) = \frac{1}{|\Theta_p|} \sum_{\theta_i \in \Theta_p} \mathrm{SE}(\theta_i), \quad p \in \{0, q, 2q, 3q, \ldots, 100\}. \tag{87}$$

Here, θ_p is the quantile value at the pth percentile and Θ_p is the set of all θ_i within the percentile range of $\max(0, p - q/2)$ and $\min(100, p + q/2)$:

$$\Theta_p = \{\theta_{\max(0,p-q/2)}, \ldots, \theta_{\min(100,p+q/2)}\}. \tag{88}$$

Note that $q = 5$ was chosen for the present analysis. In addition, the marginal reliability, root mean squared error (RMSE), and bias of θ were estimated as follows:

$$\text{Reliability}(\theta) = \frac{\sum_{i=1}^{n}(\theta_i - \bar{\theta})^2}{\sum_{i=1}^{n}(\theta_i - \bar{\theta})^2 + \sum_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2}, \tag{89}$$

$$\text{RMSE}(\theta) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2}, \tag{90}$$

$$\text{Bias}(\theta) = \frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i). \tag{91}$$

According to these four metrics, the MLE accuracy of θ seemed to be slightly higher for Sympson–Hetter, especially at the extreme ends of the ability spectrum as shown in the top plot of Fig. 2. On the other hand, the bottom plot displays the exposure rate of every item in increasing order. Perfect exposure balance would result in a completely uniform exposure rate of $L/m = 10/250 = 0.04$ for every item, which is referenced by the horizontal dashed line. The quantity L/m also happens to be the expected $\bar{R}$ when items are selected randomly and defines the lower bound of $\bar{R}$ (Chang & Zhang, 2002; Chen et al., 2003). Hence, in the current exercise, 0.04 serves as the baseline reference for computed values of average test overlap rate. Note that Sympson–Hetter and randomesque resulted in $\bar{R}^{(1)} = 0.17$ and $\bar{R}^{(2)} = 0.14$, respectively. This difference of about 0.03 was statistically significant at $\alpha = 0.01$ (i.e., $Z = 2.75 > z_{.005} = 2.58$ for a two-tailed test). Therefore, for this particular CAT design, there was objective evidence that randomesque afforded superior pool utilization or test security as quantified by the $\bar{R}$ metric. The practical significance of such an effect size, however, is more or less a value judgment based on the specific requirements of the testing program.

6. Conclusion

A mathematical explication of item exposure allowed the derivation of its distributional properties and revealed a rather straightforward relationship to test overlap. Ultimately, the asymptotic distribution of the average test overlap rate $\bar{R}$, or more precisely a linear transformation of it, was rigorously established for fixed-length CAT. This development serves to fill the theoretical gap in literature and inform operational practice on a sample statistic that is widely utilized in testing. Specifically, knowledge of $\bar{R}$'s sampling distribution enables the estimation of its standard error and construction of confidence intervals. Furthermore, differences between $\bar{R}$'s can now be evaluated statistically, permitting objective comparisons between CAT designs regarding their relative security risk or pool utilization. A realistic simulation study demonstrated a comparative evaluation of two exposure control mechanisms, which confirmed one method to be superior in controlling item exposure for a given pool under specific test conditions. Another pragmatic application would be contrasting the health of different item pool prototypes for use with a desired CAT driver, which may greatly inform the process of assembling optimal pools for a given program.

However, one caveat is that a dependency in item exposures between examinees can arise from factors other than adaptive item selection. For instance, certain exposure control mechanisms work by updating the probability of an item's selection in real time based on its previous exposures. A few examples of such methods include maximum exposure caps, item ineligibility (van der Linden & Veldkamp, 2007), and restricted maximum information (Revuelta & Ponsoda, 1998).This is not the case for common randomization techniques or methods that determine administration probabilities beforehand, such as the randomesque and Sympson–Hetter algorithms employed in the simulation study. At present, we avoided the technical complexities of dependent processes that are not ability based, but this issue certainly warrants further study.

The utility of the average test overlap rate notwithstanding, it is admittedly an incomprehensive measure of potential security risk. As a marginal mean statistic, $\bar{R}$ only captures the overall snapshot of the expected proportion of repeated items between candidate pairs. Despite a low $\bar{R}$, it is entirely possible to have wide variability in overlap across examinee pairs, perhaps as extreme as zero overlap for some and complete overlap for others. Therefore, the standard deviation of test overlap (not to be confused with the standard error of $\bar{R}$) has been suggested as a complementary measure of security (Wang, Zheng, & Chang, 2014), which has not been studied here. Moreover, a smaller $\bar{R}$ does not guarantee greater security against item compromise and preknowledge, which depends on many external factors such as the type of cheating scheme, abilities of colluders, and extent of collusion (Barrada, Olea, Ponsoda, & Abad, 2009). As with any statistical index, $\bar{R}$ should be utilized in conjunction with other relevant considerations to make prudent decisions regarding the security of a test design. Finally, acknowledging that stolen items are often shared within groups greater than two, a broader indicator of security may be a generalized version of average test overlap rate (Chang & Zhang, 2002), but its distributional properties are yet to be determined.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Baker, F. B., & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). New York, NY: Marcel Dekker.

Barrada, J., Olea, J., Ponsoda, V., & Abad, F. J. (2009). Test overlap rate and item exposure as indicators of test security in CATs. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC conference on computerized adaptive testing*.

Chang, H.-H. (2015). Psychometrics behind computerized adaptive testing. *Psychometrika*, *80*(1), 1–20.

Chang, H.-H., Qian, J., & Ying, Z. (2001). a-Stratified multistage computerized adaptive testing with b-blocking. *Applied Psychological Measurement*, *25*(4), 333–341.

Chang, H.-H., & Ying, Z. (1999). a-Stratified multistage computerized adaptive testing. *Applied Psychological Measurement*, *23*(3), 211–222.

Chang, H.-H., & Ying, Z. (2009). Nonlinear sequential designs for logistic item response theory models with applications to computerized adaptive tests. *The Annals of Statistics*, *37*(3), 1466–1488.

Chang, H.-H., & Zhang, J. (2002). Hypergeometric family and item overlap rates in computerized adaptive testing. *Psychometrika*, *67*(3), 387–398.

Chen, S.-Y., Ankenmann, R. D., & Spray, J. A. (2003). The relationship between item exposure and test overlap in computerized adaptive testing. *Journal of Educational Measurement*, *40*(2), 129–145.

Choe, E. M., Kern, J. L., & Chang, H.-H. (2018). Optimizing the use of response times for item selection in computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, *43*(2), 135–158.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York, NY: Wiley.

Han, K. T. (2016). Maximum likelihood score estimation method with fences for short-length tests and computerized adaptive tests. *Applied Psychological Measurement*, *40*(4), 289–301.

Hau, K.-T., & Chang, H.-H. (2001). Item selection in computerized adaptive testing: Should more discriminating items be used first? *Journal of Educational Measurement*, *38*(3), 249–266.

Kingsbury, G. G., & Zara, A. R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, *2*(4), 359–375.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Mislevy, R. J., & Chang, H.-H. (2000). Does adaptive testing violate local independence? *Psychometrika*, *65*(2), 149–156.

Rencher, A. C., & Schaalje, G. B. (2008). *Linear models in statistics* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

Revuelta, J., & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, *35*(4), 311–327.

Sympson, J. B., & Hetter, R. D. (1985). Controlling item-exposure rates in computerized adaptive testing. In *Proceedings of the 27th annual meeting of the Military Testing Association*. San Diego, CA: Navy Personnel Research and Development Center.

van der Linden, W. J., & Veldkamp, B. P. (2007). Conditional item-exposure control in adaptive testing using item-ineligibility probabilities. *Journal of Educational and Behavioral Statistics*, *32*(4), 398–418.

Wang, C., Zheng, Y., & Chang, H.-H. (2014). Does standard deviation matter? Using "standard deviation" to quantify security of multistage testing. *Psychometrika*, *79*(1), 154–174.

Way, W. D. (1998). Protecting the integrity of computerized testing item pools. *Educational Measurement: Issues and Practice*, *17*, 17–27.

Manuscript Received: 24 APR 2018
Final Version Received: 10 JUN 2019
Published Online Date: 1 JUL 2019