Intellectual Discourse

Volume 32 Number 2 2024

International Islamic University Malaysia
https://journals.iium.edu.my/intdiscourse/index.php/id

Intellectual Discourse

Volume 32 Number 2 2024

Intellectual Discourse is a highly respected, academic refereed journal of the International Islamic University Malaysia (IIUM). It is published twice a year by the IIUM Press, IIUM, and contains reflections, articles, research notes and review articles representing the disciplines, methods and viewpoints of the Muslim world.

Intellectual Discourse is abstracted in SCOPUS, WoS Emerging Sources Citation Index (ESCI), ProQuest, International Political Science Abstracts, Peace Research Abstracts Journal, Muslim World Book Review, Bibliography of Asian Studies, Index Islamicus, Religious and Theological Abstracts, ATLA Religion Database, MyCite, ISC and EBSCO.

ISSN 0128-4878 (Print); ISSN 2289-5639 (Online)

https://journals.iium.edu.my/intdiscourse/index.php/id
Email: intdiscourse@iium.edu.my; intdiscourse@yahoo.com

Published by:
IIUM Press, International Islamic University Malaysia
P.O. Box 10, 50728 Kuala Lumpur, Malaysia
Phone (+603) 6196-5014, Fax: (+603) 6196-6298
Website:http://iiumpress.iium.edu.my/bookshop

Intellectual Discourse
Vol. 32, No. 2, 2024

Contents

Transliteration Table: Consonants

Arabic	Roman		Arabic	Roman
ب	b		ط	ṭ
ت	t		ظ	ẓ
ث	th		ع	‘
ج	j		غ	gh
ح	ḥ		ف	f
خ	kh		ق	q
د	d		ك	k
ذ	dh		ل	l
ر	r		م	m
ز	z		ن	n
س	s		ه	h
ش	sh		و	w
ص	ṣ		ء	’
ض	ḍ		ي	y

Transliteration Table: Vowels and Diphthongs

Arabic	Roman		Arabic	Roman
◌َ	a		◌ًا، ◌ًى	an
◌ُ	u		◌ٌو	un
◌ِ	i		◌ٍي	in
◌َا، ◌ٰ، ◌َى،	ā		◌َوْ	aw
◌ُو	ū		◌َيْ	ay
◌ِي	ī		◌ُوّ	uww, ū (in final position)
			◌ِيّ	iyy, ī (in final position)

Source: ROTAS Transliteration Kit: http://rotas.iium.edu.my

Intellectual Discourse, 32:2 (2024) 585–606

ISSN 0128-4878 (Print); ISSN 2289-5639 (Online)

Development and Validation of a Tahfiz School Performance Index

Azam Othman[*]
Nik Md. Saiful Azizi Nik Abdullah[**]
Norbaiduri Ruslan[***]
Mohamad Sahari Nordin[****]
Shahrul Fhaizal Shabu[*****]

Abstract: The *Tahfiz* (Qur'an memorisation) School Performance Index was developed as a tool to measure the performance of *Tahfiz* Schools in several key areas of *Tahfiz* education and sustainable growth development. The constructs assessed in the index were developed from analyses of focus group discussions with a few groups of stakeholders. The self-developed constructs were presented to several subject matter experts in the area of Islamic education to ascertain their face and content validity. For construct validity, the index was tested on a sample of 1,384 respondents, comprising *Tahfiz* school administrators, teachers, and alumni, and analysed using the Rasch Rating Scale Model. The results generated some evidence of construct validity, although some items related to *Tahfiz* education showed some misfitting. The index could potentially be used

* Associate Professor, Department of Social Foundations and Educational Leadership, Kulliyyah of Education, International Islamic University Malaysia. Email: azam_othman@iium.edu.my. *Corresponding author*.
** Associate Professor, Department of Curriculum and Instruction, Kulliyyah of Education, International Islamic University Malaysia. Email: nikazizi@iium.edu.my
*** Assistant Professor, Department of Communication, AbdulHamid AbuSulayman Kulliyyah of Islamic Revealed Knowledge and Human Sciences, International Islamic University Malaysia. Email: norbaiduri@iium.edu.my
**** Professor, Department of Curriculum and Instruction, Kulliyyah of Education, International Islamic University Malaysia. Email: msahari@iium.edu.my
***** Independent researcher. Email:shahrulfhaizal@gmail.com

to measure and improve the quality of the *Tahfiz* education sector. The index combines aspects of *Tahfiz* education with elements of Sustainable Growth Development to offer a holistic overview of the progress and well-being of the community within a *Tahfiz* School. This study however only focused on one type of *Tahfiz* School, which is Academic *Tahfiz* Schools, a type of *Tahfiz* school that teaches academic school subjects, Islamic education subjects, and Qur'an memorisation, and excluded other types of *Tahfiz* schools such as fully traditional *Tahfiz* schools.

Keywords: *Tahfiz* Schools, *Tahfiz* education, Qur'an memorisation, *Tahfiz* School Performance Index, schools benchmarking.

Abstrak: Indeks Prestasi Sekolah Tahfiz telah dibangunkan sebagai alat untuk mengukur prestasi Sekolah Tahfiz dalam beberapa bidang utama pendidikan Tahfiz dan pembangunan pertumbuhan mampan. Konstruk yang dinilai dalam indeks dibangunkan daripada analisis perbincangan kumpulan fokus dengan beberapa kumpulan pihak berkepentingan. Konstruk-konstruk yang dibangunkan sendiri telah dibentangkan kepada beberapa pakar subjek dalam bidang pendidikan Islam untuk memastikan kesahihan wajah dan kandungannya. Untuk kesahihan konstruk, indeks telah diuji ke atas sampel 1,384 responden, yang terdiri daripada pentadbir sekolah, guru, dan alumni Tahfiz, dan dianalisis menggunakan Model Skala Penarafan Rasch. Keputusan menghasilkan beberapa bukti kesahihan konstruk, walaupun beberapa item yang berkaitan dengan pendidikan Tahfiz menunjukkan beberapa ketidaksesuaian. Indeks ini berpotensi digunakan untuk mengukur dan meningkatkan kualiti sektor pendidikan Tahfiz. Indeks ini menggabungkan aspek pendidikan Tahfiz dengan unsur-unsur Pembangunan Pertumbuhan Mampan untuk menawarkan gambaran keseluruhan holistik tentang kemajuan dan kesejahteraan komuniti dalam Sekolah Tahfiz. Kajian ini hanya memberi tumpuan kepada satu jenis Sekolah Tahfiz iaitu Sekolah Tahfiz Akademik, sejenis sekolah Tahfiz yang mengajar mata pelajaran sekolah akademik, mata pelajaran pendidikan Islam dan hafalan Al-Quran, serta tidak melibatkan jenis sekolah Tahfiz lain seperti sekolah Tahfiz tradisional sepenuhnya.

Kata kunci: Sekolah Tahfiz, Pendidikan Tahfiz, Hafalan Al-Quran, Indeks Prestasi Sekolah Tahfiz, penanda aras sekolah.

Introduction

Tahfiz schools are steadily becoming the preferred learning institution for many Muslim parents in Malaysia (Nik Abdullah et al., 2021). In recent decades, the demand for *Tahfiz* education (i.e., Qur'an memorisation)

in Malaysia has increased significantly and is expected to continue growing, especially with the establishment of academic *Tahfiz* schools (Nik Abdullah et al., 2021). These schools teach Qur'an memorisation alongside academic subjects like language and mathematics, enabling their graduates to later specialise in professional fields or Islamic disciplines such as Islamic Jurisprudence and Islamic Finance.

Tahfiz schools are a growing segment of a comprehensive Islamic education system. Thus, they require a valid and reliable empirical measurement to ensure that they fulfil the objectives of their establishment and achieve the standard of educational quality that uniquely conforms to the spirit of Islamic education. Given the unique identity of *Tahfiz* schools, a robust tool must be developed to assess their service delivery to cover critical aspects like institutional management and Islamic value inculcation. The study's concern was that most past research has focused on Islamic schools, *madrasahs*, and Indonesian *pesantrens*, using constructs not specifically developed for the *Tahfiz* school environment. For instance, Emawati (2019) explored the development of an integrated Islamic school as an emergent entity in an urban Indonesian community and the community's expectations of the school, while a study by Sa'dullah Assa'idi (2021) examined the influence of *pesantrens* (Islamic boarding schools) on the social statuses of *santris* in Indonesia. Abdullah (2019), in a similar study, looked at the school culture of several *madrasahs* in Kediri, Indonesia, and discovered the prevalence of commonly shared qualities as suggested by this paper. However, the present study differed from Abdullah (2019) as the sample did not consist specifically of *Tahfiz* schools. Although Islamic schools, *madrasahs*, and Indonesian *pesantrens* share the same aspirations as *Tahfiz* schools, there are aspects of *Tahfiz* schools that require specific attention, such as the environment for teaching Qur'an memorisation. Hence, studying the performance of *Tahfiz* schools, which are highly regulated by the Ministry of Education and other government agencies, would illuminate the significance, sustainability, and relevance of the schools in fulfilling Malaysia's national educational agenda and Islamic well-being. More importantly, examining various types of *Tahfiz* schools is crucial as it can provide insights into how their performance may vary by situational context. Therefore, this study aims to empirically assess the performance of different types of *Tahfiz* schools, addressing the existing gap in evaluating their management, operations, and overall effectiveness. Other studies have attempted to examine the performance

of Islamic schools by focusing on specific groups, such as principals (Khasanah, 2023), teachers (Istanti et al., 2020; Ahmad Saifuddin, 2020), and both (Alwi & Mumtahana, 2023), and the effectiveness of the school.

This paper discusses on the development and validation of an index that directly addresses the aspects and features specific to the *Tahfiz* school. Subsequently named the *Tahfiz* School Performance Index, it aims to empower *Tahfiz* schools to self-assess their performance as an Islamic learning institution. It can also be potentially used to benchmark different *Tahfiz* schools, allowing the institutions to learn and develop from one another, thus elevating the quality of *Tahfiz* education. The index comprises eight constructs which were developed after a series of engagements with selected groups of stakeholders, i.e., parents, school administrators, alumni, and teachers. The initial set of items aimed to gather a comprehensive view of a *Tahfiz* school, ranging from its students' needs and administrative standards to the provision of facilities and school environment. Some aspects of the Sustainable Development Goals (United Nations, 2023), introduced by the United Nations in 2015, were also incorporated into the index's development to measure students' well-being and elements of equality. Those measures aimed to address issues on the treatment and experience of specific groups, such as underprivileged students, many of whom were selected to study at *Tahfiz* schools under scholarships or sponsorship. In essence, the index was developed based on stakeholder perspectives, whose validity as a performance measure is reported in this paper.

Objectives

The study has two primary objectives: firstly, to evaluate the validity and reliability of the *Tahfiz* School Performance Scale, and secondly, to estimate the Index Score of *Tahfiz S*chool Performance. The index aims to enable *Tahfiz* schools to measure their performance in several key areas, i.e., vision, administration, facilities, and sustainable development goals. These key areas were self-developed from interviews with stakeholders and later validated by experts in Islamic education and *Tahfiz* education. Initially, 59 items encompassing the key areas of *Tahfiz* school performance were developed. Rasch modelling was used to further validate the index's psychometric properties using a sample of 1,384 respondents.

Methods

The sample consisted of 1,384 *Tahfiz* school alumni, teachers, and administrators who were selected for the study based on their experience working and studying in the *Tahfiz* school system. The study recruited 52 registered *Tahfiz* schools from different parts of Peninsular Malaysia, to which prior negotiation of access and consent was secured.

Research Instrument

Table 1 shows the constructs and items measuring the performance of *Tahfiz* schools examined in this study:

Table 1: Constructs Measured in the *Tahfiz* School Performance Index

Construct	Definition
School Vision	The impact and importance of the *Tahfiz* school, the targets set, and its graduates' potential anticipated in the long term
Education System Quality	Features that define the *Tahfiz* school as a quality educational institution in and of itself and its level of academic excellence
Administration and Support System	Aspects of the administration and support systems that are based on piety (*taqwa*), competency, and excellence
Student Facilities	Facilities that serve the needs of students of the *Tahfiz* school`
Sustainable Development Growth Elements	Three elements of Sustainable Development Growth were incorporated into the index to assess students' well-being and the school environment. Health and Safety are related to the 3rd Sustainable Development Goal, which is to "ensure healthy lives and promote well-being for all at all ages." Peaceful Society is related to the 16th Sustainable Development Goal, which is to "promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels." Quality Education, that is inclusive and equitable, relates to the 4th goal, which is to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all."

A link to a Google Form questionnaire was shared with the selected schools and later distributed to the respondents via a few social platforms, primarily WhatsApp. All data were then collected from the Google Form backend and organised for further analysis and data cleaning using SPSS. After removing outlier responses, the final sample was prepared for further analysis using the Rasch Rating Scale Model.

Data Analysis

The Rasch analysis offers evidence of the validity and reliability of an instrument and data. The model is deemed robust in establishing the psychometric properties of the instrument which, in this case, aims to scale the performance of *Tahfiz* schools. Rasch analysis enables the mapping of the estimates of item difficulties and respondents' locations, i.e., the levels of item difficulty and respondents' ability to endorse the items along a common continuum. It was observed that the continuum scales gradually changed in the degree of difficulty and endorsability, without unexpected discontinuities. In other words, the scores of the performance quality were predicated on both person and item characteristics, which were measured on the same scale.

Rasch analysis is argued to be the only measurement analytics which is invariant when the observed data fit the requirements of the model (Bond & Fox, 2007). Unlike the Classical Test Theory, the Rasch analysis acknowledges that measurement error varies across individuals; thus, the results do not depend on a particular sample of respondents. The analytics generate linear interval scores (logit scores), characterised by additive attributes. Thus, the model enables the use of summary statistics to estimate the index score. This in turn permits the tracking and comparing of results across space and time. Such flexibility and versatility make the measurement and analysis comparable to data from different studies.

In the initial analysis, the study examined the extent to which the data satisfied the requirements of the Rasch model. Subsequently, results pertaining to the measures of item and person fit, item and person reliability, response category functioning, unidimensionality, and construct validity were examined.

Analysis Procedure

To ascertain the reliability of the measures and valid inferences to be drawn from the data, the study applied the following steps:

Goodness of Fit of Measured Item: Rasch analysis provides fit statistics to assess the psychometric properties of the questionnaire. The analysis offers two statistics to evaluate the efficacy of the data, which are the fundamental requirements of the Rasch measurement model, namely (i) infit statistics, and (ii) outfit statistics. These two fit measures, based on the conventional chi-square statistic (chi-square statistic divided by its degrees of freedom), are sensitive to unexpected observations by persons and items; hence they are indicators of the good fit of the data.

Infit statistics, which is inlier-pattern-sensitive, is sensitive to unexpected patterns of responses by persons on items. The outfit statistic, on the other hand, is an outlier-sensitive fit, and it is sensitive to unexpected observations by persons on items which are relatively very easy or very hard (Wright, 1991). High infit mean-squares indicate a mismatch between respondent ability and item difficulty (a threat to validity). In contrast, high outfit mean-squares are likely due to unexpected responses by low-ability respondents.

A fit statistic of 1.50 for an item denotes 50% more variation (or "noise") than expected by the Rasch model. This means to say that the item signals an underfitting of the model. On the contrary, low mean squares suggest overfitting or redundancy. An item with a fit statistic of 0.50 suggests 50% less variation (or "overlap") than predicted; the item overfitted the model.

In general, mean-square values around 1.0 represent negligible distortion of the measurement. Infit and outfit values > 1.5 suggest a large spread in responses, while those < 0.5 represent poor variability and high redundancy. This study used the widely used rule-of-thumb to determine an acceptable mean-square value; 0.5>MnSq>1.5. Therefore, Infit and Outfit statistics smaller than 0.5 or larger than 1.5 are considered misfits. Misfitting items manifest construct-irrelevant variance and gaps along the unidimensional continuum (Baghaei, 2008). These are problematic items that may call for deletion.

Reliability and Separation Indices: Rasch model analysis produces two types of reliability Indices: Person reliability and item reliability. Person reliability is comparable to the traditional internal consistency reliability, using Cronbach's alpha. It is based on the locations of the respondents along the measurement scale. Item reliability, on the other hand, refers to the adequacy of the number of items which are included in the analysis; the cut score of the reliability index is .75 (McCreary et al., 2013).

On the other hand, Person separation refers to the replicability of the ordering of a person's ability when they answer another set of items measuring a similar construct. Hence, the index gauges the reproducibility of the relative person's location. Item separation indicates the extent to which the hierarchical sequencing of item difficulty is reproducible if the same set of items were answered by another sample of respondents.

Low item separation (< 3) normally means that the sample is not large enough to confirm the item difficulty hierarchy of the instrument. Low person separation (< 2) implies that the instrument may not be sensitive enough to differentiate between respondents of high and low endorsement abilities, in which case additional items will be required to address the issue.

Response Category Functioning: All items in the performance questionnaire (TSP) use ordinal categories of responses. Thus, it is necessary to examine how the response options were used by the samples, the results of which would indicate whether the five-point Likert scale being used was suitable. To evaluate the functioning of the 5-point response category of the TSP questionnaire, the following were observed: (a) frequency of observation in each category (i.e., ≥ 10 that shows regularity of the response distributions), (b) progression of average logit score in a linear manner, (c) mean square residual of outfit statistics (< 2.0), and (d) thresholds between every response category and its subsequent category, which should be ordered in increasing values.

Dimensionality of the Scale: The Rasch model rests on the assumption that there is only one underlying construct measured by the set of items in the questionnaire. Simply said the questionnaire measures only one factor. This one single latent or unobserved variable would substantially account for the variability of responses across items. This

means that all items in the questionnaire are expected to share one common factor, which is the performance of *Tahfiz* schools. To evaluate the tenability of the unidimensionality assumption, the study examined the Principal Component Analysis (PCA) of the residuals, which could identify if substantial subdimensions existed among the items. To decide if dimensionality holds, the variance explained by the PCA residuals needs to be at least 40% (Linacre, 2017). Additionally, the variance explained by the first principal component of the residuals should be no more than 15%. Also, the results of the PCA need to be evaluated jointly with the results of the fit statistics to identify the nature of the multidimensionality, when the data are not unidimensional (Brentari & Golia, 2007).

Another related assumption is the local independence of the items. It is one of the requirements of a quality questionnaire. The local item independence (LID) assumption requires that a respondent's response to an item does not influence his or her response to any other item. If such influence occurs, the Rasch analysis would produce indicators of the violation of this requirement.

Construct Validity and TSP Index: Rasch analysis assesses the degree to which the response pattern matches its theoretical pattern. The analysis generates a hierarchy of item difficulties, i.e., from the most difficult to the most endorseable (easy) items. Such an ordering of item difficulties, which are positioned on an equal-interval continuum, serves as a measurement ruler for the construct. This capability increases the trustworthiness of the meaning (and its interpretation) of the Rasch estimated scores. The trustworthiness of the item difficulty hierarchy constitutes the evidence of the measure's construct validity (Baghaei, 2008; Messick, 1996).

The composite score or index score summarises the prevalence and intensity of a phenomenon, which in this case is the performance quality of a *Tahfiz* school. Under the condition that it is empirically valid, the index facilitates a comparison of school performance across groups and changes across time, hence enabling the profiling of TSP values. The literature suggests that a valid and reliable index also qualifies its broad application across time and space. It should also be simple to use, easily linked to other related surveys, stable over time, and sensitive enough to detect changes. To capitalise on its potential, the development of

the TSP index score was subjected to the following 4-step procedure: (1) ensuring that items were content-valid; (2) including items that contributed to TSP's construct validity; (3) calculating a baseline TSP index score; and (4) conducting TSP index score comparisons across stakeholder groups.

Results

An initial Rasch analysis was conducted on the whole data set of 59 items from 1,384 respondents. The results revealed that 33 of the items satisfied the levels of fit deemed critical in the Rasch measurement modelling. The other 26 items did not meet the standards. Some of these poor items contained elements of social desirability, for example, the item suggesting that the "graduates of a *Tahfiz* academic system have the potential to contribute to the *ummah* and nation."

Furthermore, the results showed that a substantial proportion of the sample did not fit the measurement model. The results found that almost one-third of the sample failed to fit the Rasch requirement. The respondents had unexpectedly responded to the questionnaire items inconsistently. This means that items which were expected to be "difficult-to-agree-with" had been positively endorsed by low-ability respondents. Still, misfit occurred when some "easy-to-agree-with" items unexpectedly failed to be endorsed by high-ability respondents.

Model-Data Fit

Considering the results of the preliminary analysis, a second Rasch analysis was performed on the data with 33 good-fit items, involving a sample of 877 respondents. The final set of 33 items measuring the construct of *Tahfiz* School Performance (TSP) with the 5-point agreement scale that met the requirements of Rasch measurement is shown in Table 2.

Table 2: Item Fit

ENTRY SCORE	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASR-AL		EXACT MATCH		ITEM CODE	ITEM DESCRIPTION
			MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%		
3823	1.55	.07	1.45	7.02	1.49	6.95	.69	.75	50.8	61.7	KK3KK	Facilities for students with disabilities
3867	1.35	.07	.89	-2.35	.89	-1.68	.75	.74	66.5	62.5	TADBIR3T	Audit and accreditation systems (internal and external)
3904	1.17	.07	.77	-4.05	.77	-3.70	.76	.72	70.2	63.2	TADBIR4T	Performance indicator system (individual to school level)
3915	1.11	.07	.95	-.45	.95	-.76	.72	.72	65.7	63.3	TADBIR6T	Human resource management (hiring, promotion)
3931	1.03	.07	.88	-2.17	.88	-1.83	.73	.71	66.8	63.9	TADBIR7T	Financial management (accounting, payment)
3941	.98	.07	.86	-1.73	.86	-2.04	.72	.70	67.8	64.1	TADBIR8T	School assets and facilities management
3954	.91	.07	1.13	2.24	1.30	3.82	.66	.70	60.7	64.3	K4K	Co-curricular activities (uniform bodies, clubs)
3956	.90	.07	.99	-.18	1.03	.36	.69	.70	64.8	64.3	K7K	Study skills development (e.g. revision planning)
4002	.65	.07	.92	-1.37	.90	-1.31	.69	.68	68.4	66.2	TADBIR2T	Supervisory committee for tahfiz curriculum
4006	.63	.08	1.06	1.04	1.08	1.04	.67	.67	64.6	66.3	KK1KK	School grounds security (e.g. CCTV)
4038	.44	.08	1.21	3.26	1.57	5.58	.59	.66	59.5	67.9	K2K	Conducive environment for memorising (comfortable and clean)
4041	.43	.08	.91	-1.45	.90	-1.14	.67	.65	68.2	68.0	K8K	Leadership qualities development (e.g. trainings, camps)
4104	.03	.08	.95	-.76	.93	-.62	.63	.62	75.4	72.0	EKUITI4E	Students receive their needed attention (on an individual basis)
4109	-.01	.08	1.03	.49	.86	-1.34	.63	.61	76.2	72.3	F15F	Sickbay/treatment room
4115	-.05	.08	.93	-1.08	.92	-.68	.63	.61	75.7	72.6	KK4KK	Monitoring of negative behaviours (e.g. bullying, theft)
4115	-.05	.08	1.06	.93	.95	-.40	.60	.61	73.1	72.6	AMAN5AMA	Involvement in public services
4122	-.10	.08	1.01	.14	1.06	.54	.60	.60	75.2	73.2	EKUITI3E	Students are given equal opportunities
4141	-.24	.09	.84	-2.53	.65	-3.34	.63	.59	80.4	74.6	KK5KK	Concern for students' problems (personal, financial, etc.)

ENTRY SCORE	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASR-AL		EXACT MATCH		ITEM CODE	ITEM DESCRIPTION
			MNSQ	ZSTD	MNSQ	ZSTD	CORR	EXP	OBS	EXP		
4146	-.27	.09	1.05	.75	.94	-.46	.57	.59	70.9	75.0	K9K	Enhancing acts of worship (e.g. nafl solat and fasting)
4150	-.31	.09	1.20	2.80	1.16	1.28	.55	.58	74.8	75.3	F3F	Transportation facility (car, van, bus, etc.)
4160	-.38	.09	1.03	.46	.83	-1.40	.59	.57	79.6	76.1	F14F	Counselling room
4162	-.40	.09	1.09	1.35	.88	-.96	.57	.57	76.9	76.2	F9F	Adequate clean toilets
4169	-.46	.09	.89	-1.68	.76	-1.88	.60	.57	81.8	76.8	EKUITI5E	Underperforming students are given additional guidance
4182	-.56	.09	1.36	4.55	1.11	.82	.50	.56	71.7	77.8	W5W	Academic tahfiz could produce students who are strong believers
4195	-.68	.09	.94	-.85	.71	-2.13	.57	.54	82.2	78.9	F10F	Comfort hostel accommodation (if provided)
4198	-.70	.09	.80	-2.88	.62	-2.88	.59	.54	83.0	79.1	KK6KK	Clean school environment
4200	-.72	.10	1.06	.82	.72	-2.03	.56	.54	82.4	79.3	F5F	Shop or cooperative that caters to students' needs
4202	-.74	.10	1.07	.95	.82	-1.21	.54	.54	81.9	79.5	F4F	Library or resource centre
4212	-.83	.10	1.21	2.58	.87	-.81	.52	.53	82.6	80.3	F6F	School hall
4238	-1.10	.10	.88	-1.48	.57	-2.74	.54	.50	86.0	82.5	F8F	Mess hall or canteen
4239	-1.11	.10	1.23	2.68	.99	-.02	.45	.50	77.3	82.5	W3W	Academic tahfiz graduates are able to contribute to development
4240	-1.12	.10	1.18	2.15	.91	-.41	.47	.50	77.4	82.6	W4W	Tahfiz education could produce successful individuals
4260	-1.35	.11	1.02	.30	.68	-1.65	.49	.47	85.4	84.4	F1F	Masjid or musolla
4092.0	.00	.09	1.03	.3	.93	-.5			73.4	72.7		
121.2	.0	.01	.15	2.3	.22	2.2				6.8.29		

Table 2 displays the distributions of infit MnSq and outfit MnSq which are the indicators of fitting or misfitting items. Except for the item, *"conducive learning environment for memorisation"* (K2K), all items did not fall within the misfitting infit MnSq and outfit MnSq values. A closer look at the misfitting item found that the content of the item is essential to the measurement of *Tahfiz* School performance. Also, it did not replicate any other items and was therefore not discarded from the analysis. The values of infit MnSq of other items ranged from 0.78 (TADBIR4T) to 1.45 (item KK3KK); the estimated outfit MnSq values were between 0.57 (F8F) and 1.57 (K2K). The results suggested that the TSP items were reasonably productive.

Additionally, all items had positive correlations as measured by the point measure correlation (PTMEA CORR). The PTMEA CORR ranged between .45 (item W3W) and .76 (item TADBIR4T). The 33 questionnaire items were aligned in the same direction on the performance construct. Thus, the results supported the goodness of fit and hence, the validity of the instrument.

Reliability and Separation: To further diagnose the model-data fit, the study evaluated the reliability and separation of persons and items. The results of Rasch analysis showed that the person reliability index, which represents the reproducibility of person order, was at .79. This means that a similar ordering of the respondent ability is reproducible if they are answering another questionnaire which measures a similar construct. However, the person separation index was merely 1.96, which is just slightly lower than the acceptable cut score of 2.0. Thus, it was concluded that the TSP measure was considerably limited in its ability to locate respondents into different levels of ability to endorse.

The item reliability index, which represents the replicability of the hierarchy of item location, was .99. It would represent an almost perfect reproduction of item difficulty that could be observed if the TSP questionnaire were to be taken by another group of people. The item separation index was reasonably large, at 8.88. Hence, it was judged that the TSP measure had the property to discriminate difficulty levels among items.

Category functioning: The TSP questionnaire used five-response options, ranging from Category "1" (*Strongly Disagree*) to Category "5" (*Strongly Agree*), that the sample would select to indicate their response

to each statement in the questionnaire. To examine the practicality of using the 5-point Likert data, the study examined the functioning of the response category. Table 3 summarises the results of the response category functioning.

Table 3: Summary Statistics of Category Functioning

CATEGORY LABEL	OBSERVED SCORE	COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE
1	1	12	0	-.49	-.91	1.28	1.35	NONE	(-4.20)
2	2	166	1	.45	.22	1.16	1.18	-2.98	-2.15
3	3	1399	5	1.76	1.61	1.15	1.13	-1.23	-.15
4	4	6324	22	3.09	3.19	.97	.83	.87	2.14
5	5	21040	73	5.08	5.05	.97	.97	3.34	(4.50)

The results showed that the number of observations for Category 1 is 12, and it increased to more than 21040 observations in Category 5. The average measures increased with the category scale, indicating an appropriately behaving rating scale. The outfit MnSq of the five categories ranged from 0.83 to 1.35, suggesting that the thresholds between the categories were properly ordered. The Rasch-Andrich thresholds matched the ordering of the response categories as they increased monotonically from -2.98 (Category 1) to 3.34 (Category 5).

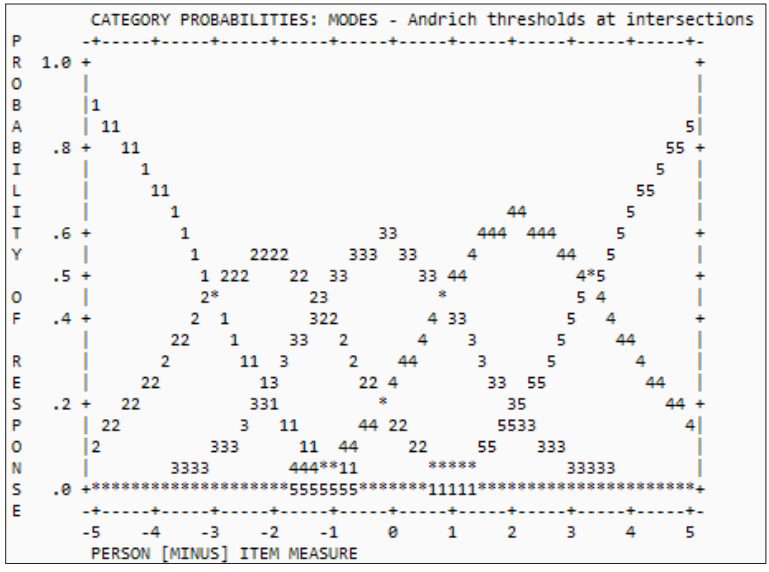

Figure 1: Rasch Model Category Probability Curves

Figure 1 displays the Rasch Category Probability Curves which supported the usefulness of the 5-point rating scale. The category widths, as shown by the differences between two successive thresholds (*), were around 2.0. Hence, the response categories were used evenly by the sample. There were no serious issues indicating that the rating scale contained overlapping categories or categories that were too far apart from each other. The 5-point TSP rating scale allowed enough information for the respondents to discriminate among response categories.

The response categories also captured a sufficient range of the TSP construct. The distance between the first category threshold, as indicated by their standard deviation (SD), ranged between -3.0 and +3.0 on a rating scale. Thus, the spread of the response categories generated a reasonably greater coverage in the measuring of the construct.

Dimensionality

To evaluate the likelihood of the variability of responses to TSP items being explained by a single underlying construct, the study examined the results of Principal Component Analysis (PCA) of the residuals. The PCA residuals showed that the Rasch dimension explained 53.5% of the variance in the data, which exceeded the 40% benchmark (Linacre, 2017). The largest secondary dimension, "the first contrast in the residuals" with an eigenvalue of 4.6 explained only 7.5% of the variance. Checked against the widely used benchmarks, the PCA of the residuals supported the unidimensionality assumption of the instrument.

However, the variance explained by the items (17.2%) was less than twice the variance explained by the first contrast. Seemingly the second dimension systematically accounted for non-random variation, which would constitute an extra dimension. The eigenvalue of the first contrast is 4.6, with a strength of about five items. Figure 2 displays the results of the unrotated principal component analysis of the standardised residuals. Letters "A, B, C,..." and "a, b, c,..." identify items with the most opposed factor loadings.

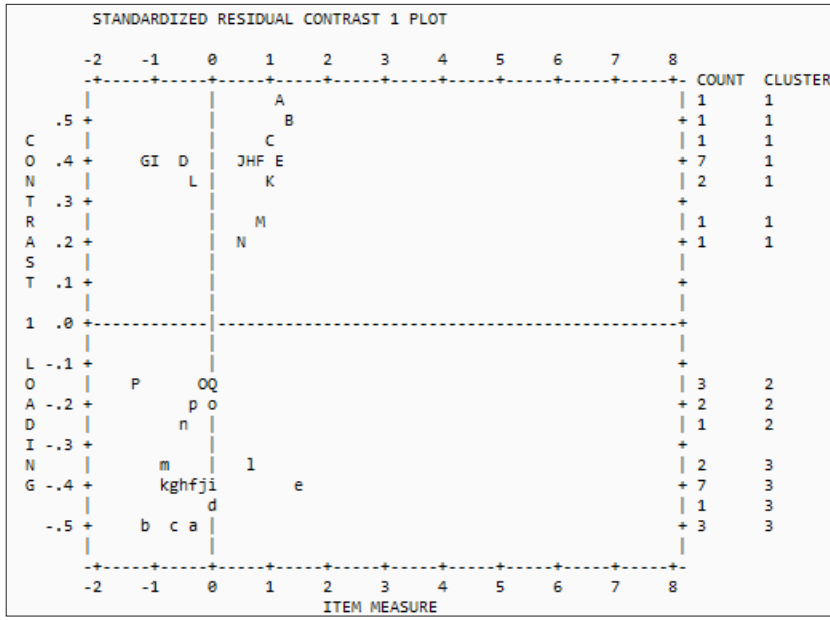

Figure 2: Plot of the Standardised Residual Contrast 1

Nevertheless, a closer inspection of the results (see Figure 2) revealed that the secondary component contained three quality performance items related to facilities. The three items (A, B, and C) loaded on an essential aspect of school performance, i.e. quality assurance. On the other hand, the three opposing items (a, b, and c) were related to school facilities, namely "adequate number of clean toilets" (a; F9f), "canteen/ dining hall" (b; F8f), and "hostel, if relevant" (c; F10f). These items were also part of the "quality school performance" construct, albeit conceptually different.

In addition, the values of the infit MnSq and outfit MnSq statistics of these items were 1.09 and 0.88, respectively, for item F9f; 0.88 and 0.57, respectively, for item F8f; 0.94 and .71, respectively, for item F10f. The statistics were within the range of fitting unidimensional items (Bond & Fox, 2013). The study decided not to discard the secondary dimension as it was reasonable to consider this component as an aspect or a strand of *Tahfiz* school performance.

The assessment of item dependency found no serious violations of the LID (local item dependency) assumption. The highest positive

residual correlation was .54, indicating the absence of local item dependency (LID) between pairs of items. The 33 items shared less than one-half of their "random" variance, suggesting that all the items were important for the unidimensional measure.

Construct Validity of the Tahfiz School Performance Measure

The validity of a measure has been classified into four types, which are predictive, concurrent, content and construct validity. The first two types are also called criterion-related validity. Of these, construct validity is the most important (Baghaei, 2008; Cronbach & Meehl, 1955; Messick, 1996). Evidence of a measure's construct validity is supported if the analysis addresses the concern, "Does the item difficulty hierarchy make sense?"

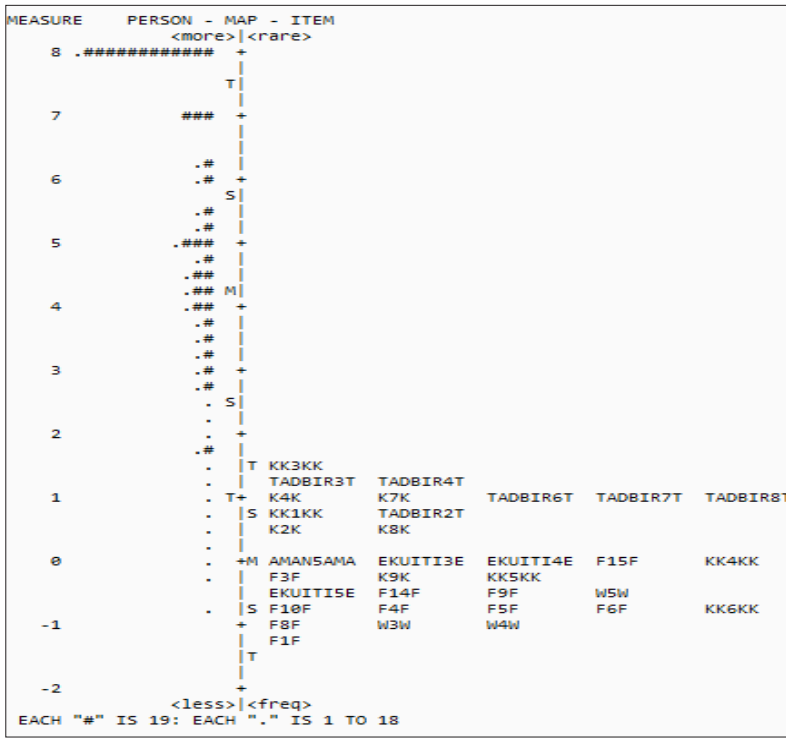

Figure 3: Person-Item Map

As demonstrated earlier, the construct validity of the TSP is supported by several benchmarks of the Rasch model, which are summarised as follows:

1. All point measure correlations were positive with a minimum value of .49
2. The estimated item reliability was .99
3. The item separation index was 8.88 and the item reliability was .99
4. Reasonableness of the assumption of unidimensionality
5. Tenability of the assumption for the absence of local item dependency (LID)

The Wright map (see Figure 3) allows for a visual comparison of the Rasch analysis' predicted ordering of item difficulty against the actual order of item difficulty in a data set. It tests the Rasch model-data alignment to substantiate the measure's construct validity (Boone & Staver, 2014). The map facilitates an efficient assessment of the validity.

Figure 3 depicts the hypothetical unidimensional *Tahfiz* school performance construct that the study aimed to measure using the 33-item questionnaire. The item at the top of the line (i.e., the existence of facilities for students with disabilities--item KK3KK) was least agreeable to the sample. As expected, it was a difficult item to endorse, given the context of the study. The persons (symbol, #) located at the top of the line were more able to agree. At the bottom of the line, the items became easier (the easiest being praying space; F1F), and the persons became less able to endorse.

The inspection of the Wright map suggested the hierarchy of the estimated item difficulty matched its theoretical ordering. Moreover, the map showed the absence of huge gaps between the items, indicating that the construct was not underrepresented. The results, therefore, offered evidence of construct validity for the TSP measure.

However, the Wright map indicates that there is no item that targets the respondents at the higher end of the continuum. In other words, the targeting between the items and sample (i.e., whether the item difficulty range matches the sample ability range) was weak. The map shows that a big majority of the respondents were able to easily endorse the 33-item instrument. Therefore, the current instrument developed may not necessarily address all possible items that measure the constructs of performance of Tahfiz schools. Future studies should address this concern, namely by looking into including items that are difficult to endorse

Estimates of Index Scores

The Rasch analysis estimated the respondents' logit scores, which exhibited additive properties. The sample of 877 respondents' scores ranged from -0.82 to 8.34, with a mean score of 5.33 and a standard deviation of 2.30. The scores were then extrapolated to be positively distributed to ease interpretations. The eyeball inspection found that the person data was normally distributed; skewness and kurtosis statistics were estimated at 1.5 (Tabachnick & Fidell, 2013).

Figure 4 illustrates the performance index score of the schools. The distributional analysis of the students' scores showed that, on a scale of 0-10, the composite score of *Tahfiz* schools' performance was 6.32, with a standard deviation of 2.29. The analysis also revealed that the teachers were slightly more inclined to positively endorse the quality of *Tahfiz* schools (Index Score = 6.50, SD = 2.25). The two administrative personnel and alumni respectively scored 6.33 (SD = 2.14) and 6.08 (SD = 2.36).

The findings as shown in Figure 4 indicate that the *Tahfiz* School Performance Index is highest among the teachers. In other words, among the sample, teachers tend to give the highest index of TSP. The next highest index was given by administrators, followed by the alumni.

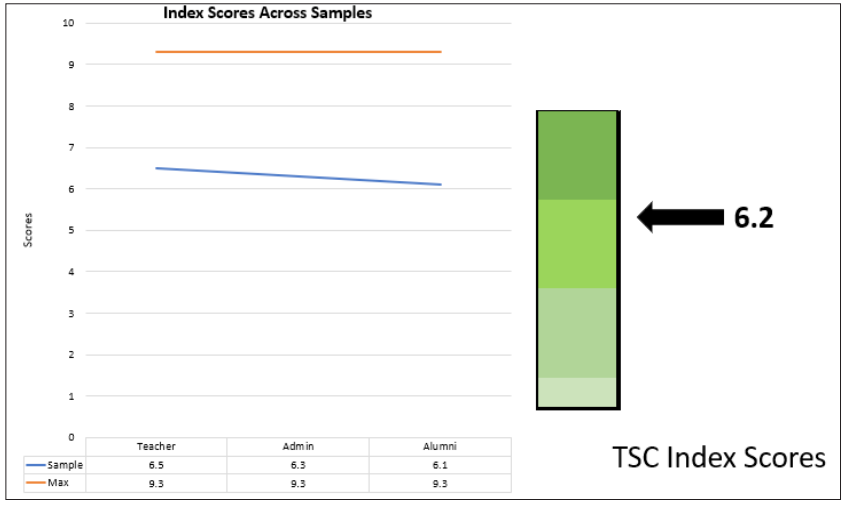

Figure 4. The *Tahfiz* School Performance Index

Conclusion and Recommendations

The *Tahfiz* School Performance Index is an attempt to improve the quality of education at *Tahfiz* schools by having a standardised measurement that includes the key areas that are important to their performance. Using the Rasch Rating Scale Model, this paper has demonstrated the validity and reliability of the index, with specific evidence of construct validity. At the same time, further work is needed to address some issues related to the index, such as the ease of endorsement of the index's items, as indicated in the Wright map. In the present analysis, certain items related to *Tahfiz* education were misfitting and had to be excluded. Further examination is needed to determine the validity of these items, which would improve the index as an indicator of *Tahfiz* School performance.

The index can be used by three stakeholders i.e., teachers, students, and administrators. Teachers and students can use this index to inform the administration of their school's strengths and weaknesses, while administrators can use the index to self-assess their performance and benchmark it against that of others over time. They may conduct the benchmarking as an annual exercise. This index offers *Tahfiz* schools a comprehensive measure of performance that not only addresses the academic aspect but also the administrative, well-being, and facilities aspects.

As the demand for *Tahfiz* education increases, performance measures are important to ensure that *Tahfiz* schools can offer quality education. For governmental bodies like the Ministry of Education, this index can be used to monitor the progress and performance of the *Tahfiz* education sector. Empirical knowledge of the status of *Tahfiz* schools, as well as their conditions, strengths, and weaknesses, would enable educational authorities to develop strategies that would continuously enhance the quality of *Tahfiz* education for the public.

References

Ahmad Saifuddin, W. (2020). The influence of transformational leadership, job satisfaction and organizational citizenship behavior on the performance of Islamic school teachers. *Systematic Reviews in Pharmacy*, 11(7). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3714043 (accessed 18 January 2023).

Alwi, M. & Mumtahana, L. (2023). The principal's strategy in improving the quality of teacher performance in the learning process in Islamic elementary schools. *Kharisma: Jurnal Administrasi dan Manajemen Pendidikan*, 2(1), 66-78. Available at doi:10.2991/jnmp.2006.13.4.1 (accessed 18 January 2023).

Baghaei, P. (2008). The Rasch model as a construct validation tool. *Rasch Measurement*, 22(1), 1145-1154.

Brentari, E. & Golia, S. (2007). Unidimensionality in the Rasch model: how to detect and interpret. *Statistica*, 67, 253-261.

Bond, T. G. & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences.* Psychology Press: New York.

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52*,* 281-302.

Emawati. (2019). Explore the development of an integrated Islamic school as an emergent property, branding, and expectations of Islamic schools. *Advances in Social Science, Education and Humanities Research,* 408, 166-171.

Istanti, E., Soeherman, A. D. G., Budianto, F., Noviandari, I. & Sanusi, R. (2020). The influences of motivation, work milieu, and organizational commitment on teacher performance in MTS Negeri 4 (Public Islamic School), Surabaya East Java. Available at https://www.ijicc.net/images/vol_13/Iss_2/SC65_Istanti_2020_E_R.pdf (accessed 18 January 2023).

Khasanah, F., Zainuddin, M., Ramli, A., Susanto, P. & Sesario, R. (2023). The analysis role of social skills and principal's performance on school's culture of private Islamic school. *Journal on Education*, 5(4), 12980-12985. Available at https://jonedu.org/index.php/joe/article/view/2290 (accessed 18 January 2023).

Linacre, J. (2017). *Winsteps® (Version 3.93.0)* [Computer Software]. Available at http://www.winsteps.com (accessed 18 January 2023).

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.

McCreary, L. L., Conrad, K. M., Conrad, K. J., Scott, C. K., Funk, R. R. & Dennis, M. L. (2013). Using the Rasch measurement model in psychometric analysis of the family effectiveness measure. *Nursing Research*, 62(3), 149-159.

Mukhammad Abdullah. (2019). School culture to serve performance of Madrasah in Indonesia. *Qudus International Journal of Islamic Studies*, 7(1), 71-100.

Nik Abdullah, N.M.S.A., Muhammad Isa, R.A., and Mohd Sabbri, F.S. (2021), *Tahfiz* students' experiences in memorizing the Qur'an: unveiling their

motivating factors and challenges. *IIUM Journal of Educational Studies, 9(2),* 42-63.

Sa'dullah Assa'idi. (2021). The growth of pesantren in Indonesia as the Islamic venue and social class status of santri. *Eurasian Journal of Educational Research,* 93, 425-440.

Tabachnick, B. & Fidell, L. S. (2013). *Using Multivariate Statistics*, Pearson: New York.

United Nations. (2023). *The 17 goals: Sustainable Development.* Available at https://sdgs.un.org/goals#history (accessed 18 January 2023).

Wright, B. (1991), Diagnosing misfit. *Rasch Measurement Transactions*, 5(2).

GUIDELINES FOR AUTHORS

Intellectual Discourse is an academic, refereed journal, published twice a year. Four types of contributions are considered for publication in this journal: major articles reporting findings of original research; review articles synthesising important deliberations related to disciplines within the domain of Islamic sciences; short research notes or communications, containing original ideas or discussions on vital issues of contemporary concern, and book reviews; and brief reader comments, or statements of divergent viewpoints.

To submit manuscript, go to http://www.iium.edu.my/intdiscourse

The manuscript submitted to Intellectual Discourse should not have been published elsewhere, and should not be under consideration by other publications. This must be stated in the covering letter.

1. Original research and review articles should be 5,000-8,000 words while research notes 3,000-4,000 words, accompanied by an abstract of 100-150 words. Book review should be 1,000-1,500 words.
2. Manuscripts should be double-spaced with a 1-inch (2.5 cm) margins. Use 12-point Times New Roman font.
3. Manuscripts should adhere to the *American Psychological Association* (APA) style*,* latest edition.
4. The title should be as concise as possible and should appear on a separate sheet together with name(s) of the author(s), affiliation(s), and the complete postal address of the institute(s).
5. A short running title of not more than 40 characters should also be included.
6. Headings and sub-headings of different sections should be clearly indicated.
7. References should be alphabetically ordered. Some examples are given below:

Book
In-text citations:
Al-Faruqi & al-Faruqi (1986)

Reference:
Al-Faruqi, I. R., & al-Faruqi, L. L. (1986). *The cultural atlas of Islam*. New York: Macmillan Publishing Company.

Chapter in a Book
In-text:
Alias (2009)

Reference:
Alias, A. (2009). Human nature. In N. M. Noor (Ed.), *Human nature from an Islamic perspective: A guide to teaching and learning* (pp.79-117). Kuala Lumpur: IIUM Press.

Journal Article
In-text:
Chapra (2002)

Reference:
Chapra, M. U. (2002). Islam and the international debt problem. *Journal of Islamic Studies, 10*, 214-232.

The Qur'ān
In-text:
(i) direct quotation, write as 30:36
(ii) indirect quotation, write as Qur'ān, 30:36

Reference:
The glorious Qur'ān. Translation and commentary by A. Yusuf Ali (1977). US: American Trust Publications.

Ḥadīth
In-text:
(i) Al-Bukhārī, 88:204 (where 88 is the book number, 204 is the ḥadīth number)
(ii) Ibn Hanbal, vol. 1, p. 1

Reference:
(i) Al-Bukhārī, M. (1981). *Ṣaḥīḥ al-Bukhārī*. Beirut: Dār al-Fikr.
(ii) Ibn Ḥanbal, A. (1982). *Musnad Aḥmad Ibn Ḥanbal*. Istanbul: Cagri Yayinlari.

The Bible
In-text:
Matthew 12:31-32

Reference:
The new Oxford annonated Bible. (2007). Oxford: Oxford University Press.

Transliteration of Arabic words should follow the style indicated in ROTAS Transliteration Kit as detailed on its website (http://rotas.iium.edu.my/?Table_of_Transliteration), which is a slight modification of ALA-LC (Library of Congress and the American Library Association) transliteration scheme. Transliteration of Persian, Urdu, Turkish and other scripts should follow ALA-LC scheme.

IIUM Press (Marketing Unit)
Research Management Centre
International Islamic University Malaysia
P.O. Box 10, 50728 Kuala Lumpur, Malaysia
Phone (+603) 6196-5014, Fax: (+603) 6196-4862
E-mail: intdiscourse@iium.edu.my; intdiscourse@yahoo.com.
Website: http://iiumpress.iium.edu.my/bookshop

In This Issue

ISSN 0128-4878 (Print)

ISSN 2289-5639 (Online)

ISSN 0128-4878
2289 5639