

# Proposal to encode an Arabic tail character used for abbreviation

Roozbeh Pournader (WhatsApp) and Borna Izadpanah (University of Reading)

May 1, 2020


## Background

One of the authors, Borna Izadpanah, is a researcher into the history of Arabic script printing and typography. The other author, Roozbeh Pournader, “discovered” the character proposed here in a presentation given by Borna Izadpanah at IStype 2019. Further research led to finding several more examples of the proposed character.

So far, this character has been observed in several of the publications by the first printing press in Tehran, known as *Chap-e Mo'tamedī* or *Mo'tamedī Imprint*, established in 1820s. The metal types with which these publications were printed were most likely designed in Iran by Iranians. These types were based on the Persian flavor of the Naskh style. Encoding the character would help in the digitization efforts of early Persian movable type printing in Iran.

## Proposal

Encode the following character in the Unicode Standard:

Glyph	Codepoint	Name
	088E	ARABIC VERTICAL TAIL

The proposed character properties follow:

### UnicodeData.txt

088E; ARABIC VERTICAL TAIL;Lo;0;AL;;;;;N;;;;;

### ArabicShaping.txt

088E; VERTICAL TAIL; R; VERTICAL TAIL

All other properties should be similar to U+08AC ARABIC LETTER ROHINGYA YEH.

## Suggestion for NamesLists.txt

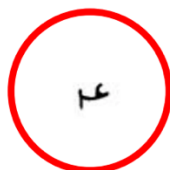
088E ARABIC VERTICAL TAIL

\* Used an abbreviation marker in early Persian movable type printing in Iran; only observed in final form

### Analysis

The character is similar in function to an abbreviation period in English. It denotes that the rest of the word or phrase is missing, and that the reader is expected to figure out the missing part from the context. In the examples of usage provided, we have mentioned the full phrase in the captions.

This abbreviation marker should predate Persian typography, since it also exists in modern Urdu honorifics which don't seem to have a clear relation to early Persian typography. The Urdu honorifics have moved above the baseline though, as can be seen in already encoded characters such as U+0611 ARABIC SIGN ALAYHE ASSALLAM, seen here in a sample from L2/01-426:



peace and blessings be upon him!

sign placed over the names of other Prophets

It stands for عَلَيْهِ السَّلَام /ale[h]yssa

"upon him be peace!" E. g. عيسى

We expect that over time, more examples of the proposed character will be found, including inline usage in older Urdu texts.

Note the visible tooth after the *ain* in the sample above, which signifies that this is not a cut isolated *ain* (or another kind of modified *ain*), but an initial *ain* followed by the abbreviation marker. This is in contrast with the cut *ain* seen in the Urdu equivalent of Anno Domini, which looks like a hamza to the untrained eye, but is indeed a cut isolated *ain*, short for عيسوى ("of Jesus"). Here's an example from The Unicode Standard, Version 13.0, page 372, Figure 9-6:



The proposed character has some similarity with U+FE73 ARABIC TAIL FRAGMENT (see L2/01-095). But that character came from legacy character sets that did not have automatic

shaping and is only expected to occur next to other presentation forms that don't have automatic shaping either. Our proposed character should instead be encoded as a normal shaping character, so that it can come after normal Arabic letters and affect their shapes.

We consider the character proposed here to be a separate character with clear semantic and orthographic distinctions and not a modification of the letter before it. We believe it is a letter-like shaping character which has been used productively in combinations with other letters.

Theoretically, instead of one character, more characters could be encoded: a *sad* followed by a tail, an *ain* followed by a tail, etc. But considering that:

1. The proposed character is a separate piece of writing (and a separate piece of type, as can be seen in Figure 5) with no semantic relation to its previous letter except for happening to occur after it; and
2. We expect examples of the character to be found after other letters, such as after *dad* (as used in U+0613 ARABIC SIGN RADI ALLAHOU ANHU), and we don't want to turn this into a "Where's Waldo?" chase across hard-to-find early Persian books and manuscripts, with proposal after proposal to come when we find new examples;

We think it serves the Unicode Standard better if it is encoded as a single character.

Another objection may be that this may not be a plain text character, but an advanced typographic feature, comparable to a swash. Considering that the character has a very clear semantic value and removing it from the plain text representation results in unreadable and meaningless words such as *حقتع*, as well as noticing that the sample texts otherwise contain no typographic or aesthetic feature with semantic value, we reject that analysis.

We have only observed the proposed character in its final form. We don't know if it ever occurs in isolated form. In this way, it's similar to U+08AC ARABIC LETTER ROHINGYA YEH (see L2/10-288R). If a design for the not-yet-observed isolated form is desired, it could look like the final form with no right-side connection.

## Samples of usage

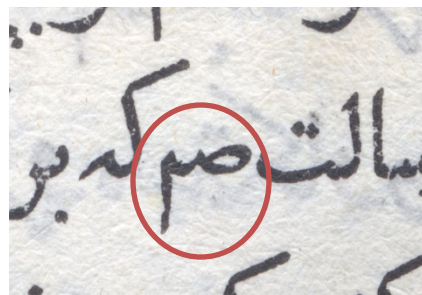


Figure 1. The proposed character occurring after *sad*, following a reference to the prophet Muhammad to represent the honorific *صلی الله علیه* or one of its variations such as *صلی الله علیه و آله* or *صلی الله علیه و سلم* or *صلی الله علیه و سلم و آله*. From Majlesi 1825, page 71.

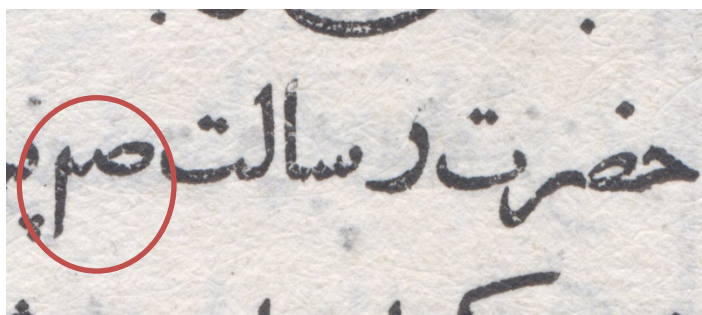


Figure 2. The proposed character occurring after *sad*, following a reference to the prophet Muhammad, to represent the honorific *عليه* or its variations. From Majlesi 1825, page 71.

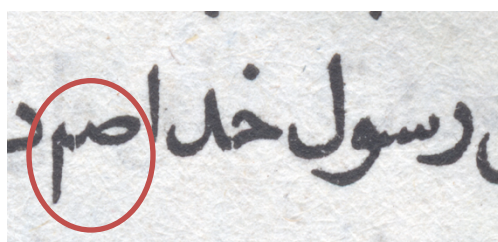


Figure 3. The proposed character occurring after *sad*, following a reference to the prophet Muhammad, to represent the honorific *عليه* or its variations. From Majlesi 1825, page 56.

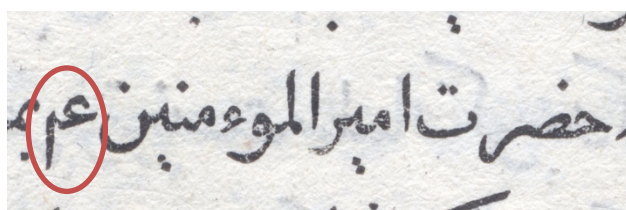


Figure 4. The proposed character occurring after *ain*, following a reference to Ali, to represent the honorific *عليه السلام*. From Majlesi 1825, page 71.

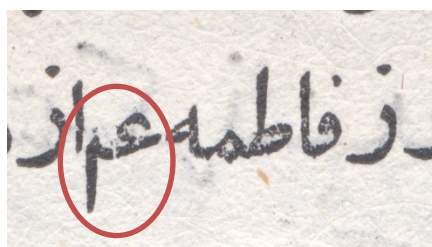


Figure 5. The proposed character occurring after *ain*, following the name of Fatimah, to represent the honorific *عليها السلام*. From Majlesi 1825, page 72. Note the gap between *ain* and the tail, showing that this was most probably a different piece of type.

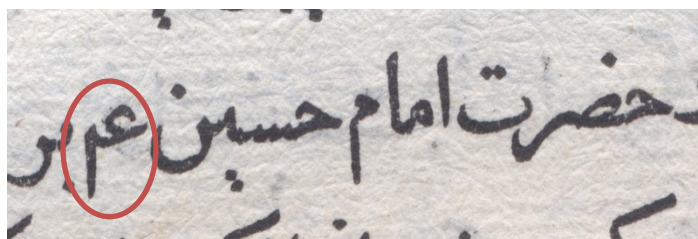


Figure 6. The proposed character occurring after *ain*, following the name of Husayn ibn Ali, to represent the honorific علیه السلام. From Majlesi 1825, page 53.

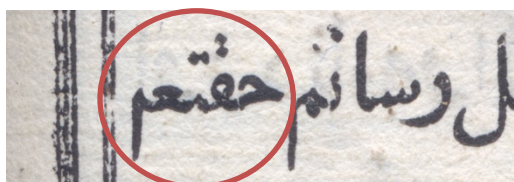


Figure 7. The proposed character occurring at the end of a word-like phrase, *hah-qaf-teh-ain*, as an abbreviation for the phrase حق تعالی which is a name of God. From Majlesi 1825, page 72.

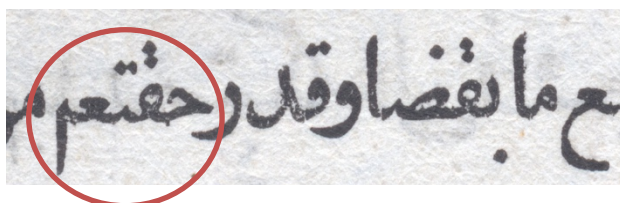


Figure 8. The proposed character occurring at the end of a word-like phrase, *hah-qaf-teh-ain*, as an abbreviation for the phrase حق تعالی. From Majlesi 1825, page 56.

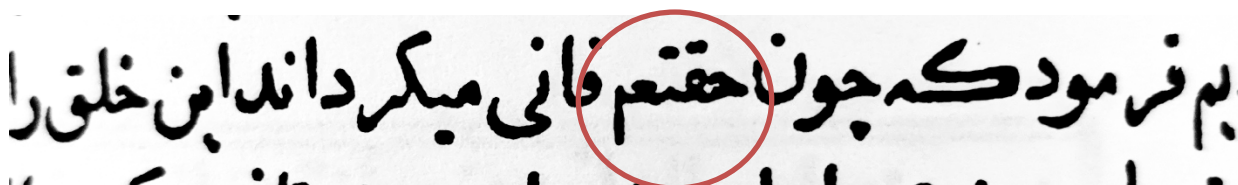


Figure 9. The proposed character occurring at the end of a word-like phrase, *hah-qaf-teh-ain*, as an abbreviation for the phrase حق تعالی. From Majlesi 1843, last page. Reproduced in Babazadeh 1999, page 215.

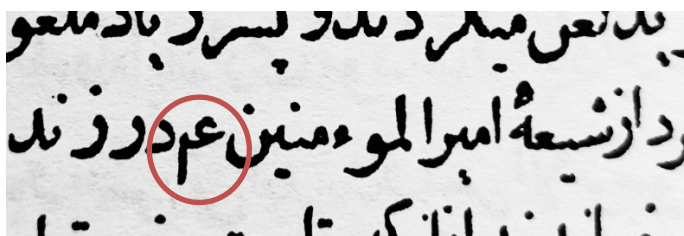


Figure 10. The proposed character occurring after *ain*, following a reference to Ali, to represent the honorific علیه السلام. From Va'ez Heravi 1845, page 37. Reproduced in Babazadeh 1999, page 242.



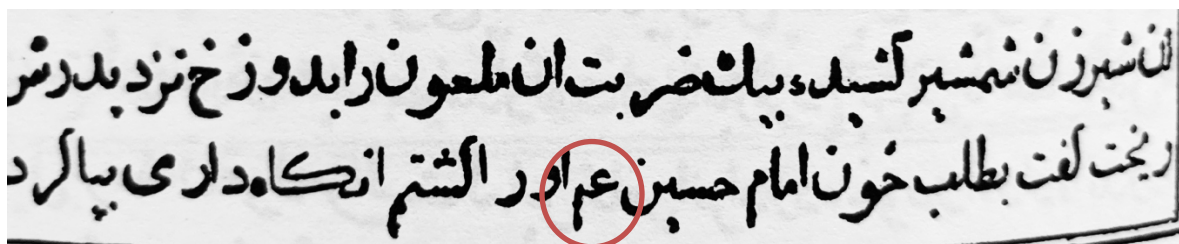


Figure 11. The proposed character occurring after *ain*, following the name of Husayn ibn Ali, to represent the honorific علیه السلام. From Va'ez Heravi 1845, page 168. Reproduced in Babazadeh 1999, page 243.

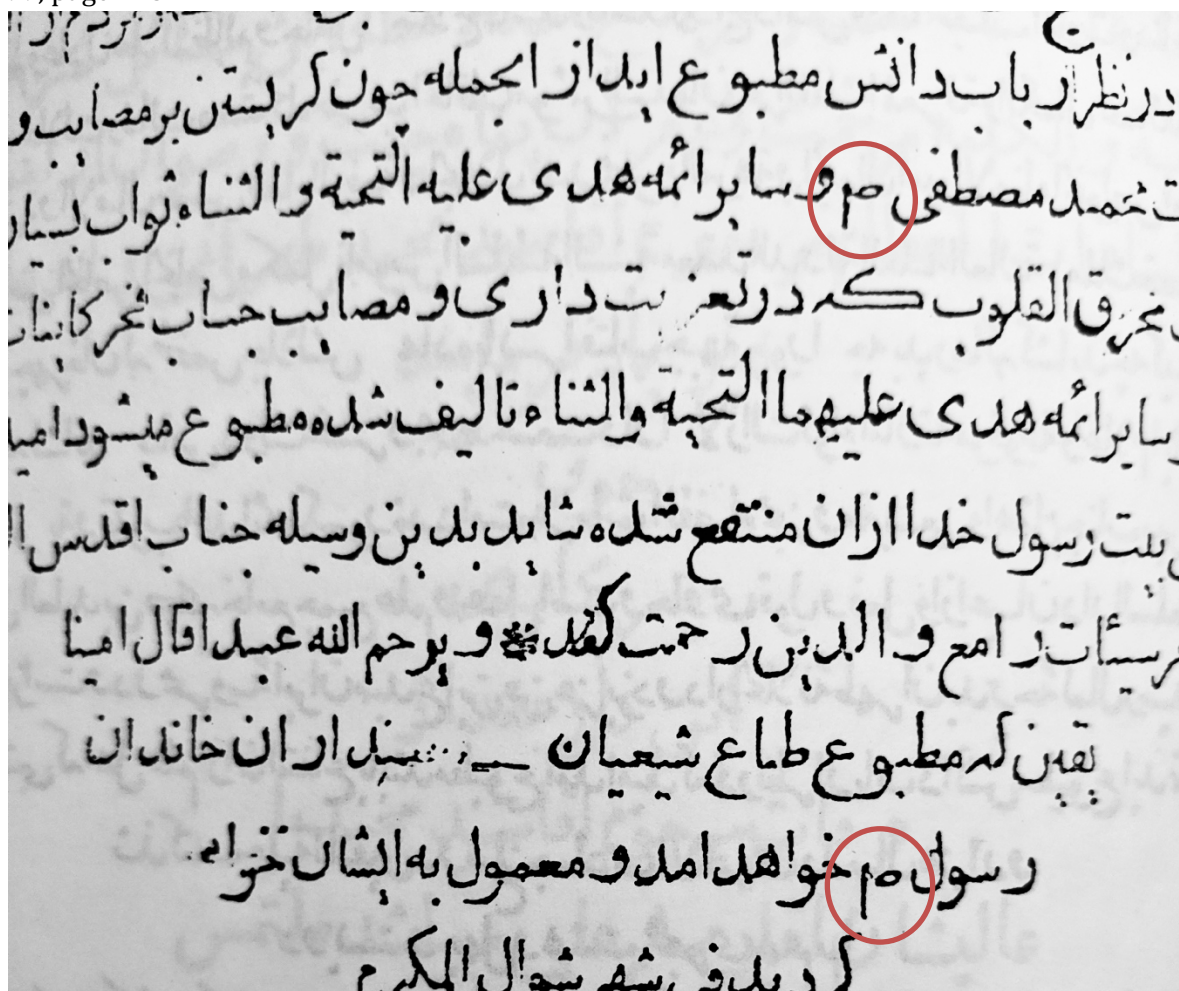


Figure 12. The proposed character occurring twice after *sad*, following references to the prophet Muhammad, to represent the honorific صلی الله علیه or its variations. From Naraghi 1832, last page. Reproduced in Babazadeh 1999, page 231. This may be a ligated form, since the *sads* appear to have lost their teeth.

میباید پس فرموده \* و باز فرموده است که و انجان خلقناه من قبل من نار السموم \*  
 خواهد بود در زمین و خواهند بود در دنیا بمنزله پیغمبران و انجم \* و اباز اینها و انجم ممکنست که ادم کشتن  
 بناحق و زنا کردن صادر شود با نیکدانی که خدا هرگز زمین و از پیغمبری با امامی از فرزندان ادم خالی  
 نداشته است با نیکدانی که خدا میفرماید که نفرستادیم پیش از تو یعنی بسوی خلق مکر مردانی چند  
 که وحی میفرستادیم بسوی ایشان از اهل شهرهای پس این دلیلست بر آنکه ملئکه را بسوی زمین  
 نفرستاده است که پیشوایان و حکام باشند بلکه ایشان را بسوی پیغمبران خود فرستاده است پس راویان  
 عرص کردند که پس بنا بر این شیطان نیز میباید ملئک نباشد فرمود که او نیز ملئک نبود بلکه از جن بود چنانچه  
 حق تعالی فرموده است که \* انه كان من الجن \* و باز فرموده است که و انجان خلقناه من قبل من نار السموم \*  
 و بدستیکه خبر داد مرا پدرم از جدم از حضرت امام و خاتم الانبیا از پدرانش از رسول خدا صلی الله علیه  
 و آله که آنحضرت فرمود که حق تعالی اختیار کرد از جمیع عالمیان محمد و آل محمد را و اختیار کرد پیغمبران را و  
 اختیار کرد ملئکه مقربان را و اختیار نکرد ایشان را مگر برای آنکه میدانست که کاری نخواهند کرد که  
 از ولایت و دوستی خدا بیرون روند و از عصمت الهی ببری شوند و ختم شوند با کس و همی که مستحق  
 عذاب خدا گردیده اند پس راویان گفتند که بما روايت رسیده است که چون حضرت رسول صلی الله علیه  
 و آله فرمود بر حضرت امیرالمومنین ع با مامت و عرض کرد خداوند عالمیان و ولایت آنحضرت را بر ملئکه پس  
 گروه بسیاری قبول ولایت آنحضرت نکردند و خدا ایشان را مسخ کرد بصورت و ذغابی فرمود  
 معاذ الله این حدیث را بر ملا روغ بسته اند و ملئکه رسولان خدا نبند و چنانچه بر پیغمبران  
 خدا انفراد نیست بر ایشان نیز و انست و شان ملئکه عظیم است و مرتبه ایشان جلیل  
 است و از امثال این امور مترجمند و باین جا منتهی شد آنچه از تفسیر امام ع نال  
 گردید و ساد احوال ملئکه و شان ع و انان از کلام و حواله و اح

Figure 13. The proposed character occurring eight times on a single page, after *sad* and *ain*, as well as the word-like phrase, *hah-qaf-teh-ain*. From Majlesi 1845, last page. Reproduced in Babazadeh 1999, page 246.

## Bibliography

1. Shahla Babazadeh. 1999. *The History of Printing in Iran (Ta'rikh-e Chaap dar Iraan)*. Tehran: Tahoori. ISBN 964-6414-08-7.
2. IBM Egypt and the Unicode Consortium. 2001. "Proposal to add "Arabic Tail Fragment" character." UTC Document Register L2/01-095, The Unicode Consortium.  
<https://www.unicode.org/L2/L2001/01095-N2322-ArabicTail.pdf>

3. Borna Izadpanah. 2019. "The emergence of printing in Qajar Iran." Presented June 15, 2019 at Istanbul Typography Conference (ISType). Video posted to YouTube on Sep 8, 2019. <https://youtu.be/n67bRO3PYBs>
4. Jonathan Kew. 2001. "Proposal to add Arabic-script honorifics and other marks: Appendix". UTC Document Register L2/01-425, The Unicode Consortium. [https://unicode.org/L2/L2001/01426-arabic\\_marks\\_examples.pdf](https://unicode.org/L2/L2001/01426-arabic_marks_examples.pdf)
5. Mohammad-Bagher Majlesi. 1825 (1240 AH). *Jalaa' al-'Ayun*. Tehran: Mo'tamedī Imprint. From the personal collection of Borna Izadpanah.
6. Mohammad-Bagher Majlesi. 1843 (1259 AH). *Haqq ol-Yaqin*. Tehran: Zeynolabedin Tabrizi. Reproduced in Babazadeh 1999.
7. Mohammad-Bagher Majlesi. 1845 (1261 AH). *Hayaat ol-Qolub*. Tehran: Mohammad Esmail, Mo'tamedī Imprint. Reproduced in Babazadeh 1999.
8. Mehdi Naraghi. 1832 (1247 AH). *Moharreq al-Qolub*. Tehran: Zeynolabedin Tabrizi of Mo'tamedī Imprint. Reproduced in Babazadeh 1999.
9. Lorna A. Priest and Martin Hosken. 2010. "Proposal to add Arabic script characters for African and Asian languages." UTC Document Register L2/10-288R, The Unicode Consortium. <https://unicode.org/L2/L2010/10288r-arabic-proposal.pdf>
10. The Unicode Consortium. 2020. The Unicode Standard. Version 13.0 — Core Specification. Mountain View, CA: The Unicode Consortium. ISBN 978-1-936213-26-9. <https://www.unicode.org/versions/Unicode13.0.0/UnicodeStandard-13.0.pdf>
11. Ata'ollah Va'ez Heravi. 1845 (1261 AH). *Mokhtaarnaameh = Rowzat ol-Mojaahedin*. Tehran: Abdolkarim. Reproduced in Babazadeh 1999.



ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646

**A. Administrative**

1. Title: ***Proposal to encode an Arabic tail character used for abbreviation***
2. Requester's name: **Roozbeh Pournader and Borna Izadpanah**
3. Requester Type: **Expert Contribution**
4. Submission date: **May 1, 2020**
5. Requester's reference, if applicable: **N/A**
6. Choose one of the following:  
    This is a complete proposal: **Yes**  
    (or) More information will be provided later: **No**

**B. Technical – General**

1. Choose one of the following:
  - a. This proposal is for a new script (set of characters): **No**  
    Proposed name of script: **N/A**
  - b. The proposal is for addition of character(s) to an existing block: **Yes**  
    Name of existing block: **Arabic Extended-B**
2. Number of characters in proposal: **1**
3. Proposed category: **D-Attested Extinct**
4. Is a repertoire including character names provided? **Yes**
  - a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? **Yes**
  - b. Are the character shapes attached in a legible form suitable for review? **Yes**
5. Fonts related:
  - a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? **Borna Izadpanah**
  - b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): **Borna Izadpanah**
6. References:
  - a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? **No**
  - b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? **Yes.**
7. Special encoding issues:  
    Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please endorse information)? **Yes.**

8. Additional information:  
Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation

behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

### C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? **No**  
If YES explain: **N/A**
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? **Yes**  
If YES, with whom? **The second author is part of the user community, an expert researcher studying the history of the Arabic script printing and typography**  
If YES, available relevant documents: **N/A**
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? **Yes**  
Reference:
4. The context of use for the proposed characters (type of use; common or rare): **Rare**  
Reference:
5. Are the proposed characters in current use by the user community? **No**  
If YES, where? **N/A**  
Reference:
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? **Yes**  
If YES, is a rationale provided? **Yes. Needs to be next to similar characters.**  
If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? **N/A**
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? **No**  
If YES, is a rationale for its inclusion provided? **N/A**  
If YES, reference: **N/A**
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? **No**  
If YES, is a rationale for its inclusion provided? **N/A**  
If YES, reference: **N/A**
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? **Yes**  
If YES, is a rationale for its inclusion provided? **Yes. See proposal.**  
If YES, reference: **See proposal.**
11. Does the proposal include use of combining characters and/or use of composite sequences?  
**No**  
If YES, is a rationale for such use provided? **N/A**  
If YES, reference: **N/A**  
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? **N/A**  
If YES, reference: **N/A**

12. Does the proposal contain characters with any special properties such as control function or similar semantics? **No**

If YES, describe in detail (include attachment if necessary): **N/A**

13. Does the proposal contain any Ideographic compatibility characters? **No**

If YES, are the equivalent corresponding unified ideographic characters identified? **N/A**

If YES, reference: **N/A**