Table 6. Results of "hometown" test for ST4

Tape A:

Question	Score	Question	Score	Question	Score	Question	Score
1	80	5	100	9	100	13	100
2	100	6	90	10	100	14	85
3	100	7	100	11	100		
4	95	8	90	12	90		

Tape B:

Question	Score	Question	Score	Question	Score	Question	Score
1	100	5	100	9	85	13	100
2	100	6	90	10	100	14	90
3	100	7	85	11	100		
4	100	8	80	12	100		

3. Problems Encountered in the Development of the Test

As mentioned in §1.4, texts that require higher levels of ability in Standard Thai were specifically used with the aim of making the test more suitable for bilingualism testing. Attempting to develop a test using non-narrative and formal-language texts presented numerous difficulties, however. The first problem to be addressed was that of how to collect texts: As mentioned previously, RTTs require oral texts, but formal texts are more usually written rather than spoken. I have assumed that this is not, in principle, a problem since news broadcasts are prepared in writing but are delivered orally, and since each of the texts that originated in written form, ST1–3, are very similar in style to stories for news broadcasts. There was an additional complication, however, in the need to have texts of roughly 2–3 minutes in length. This required that the written sources used for ST1–3 be adapted to shorten them which, in turn, introduced the possibility that the resulting texts were less coherent, less natural, or that crucial pieces of information might have been removed making the resulting text difficult to understand. Hence an additional step of checking and revision was required to ensure that the texts were good. Portions of ST2 in particular were difficult to understand in the first draft, and a few rounds of checking and revision were required.

Even after the original texts were adapted for the purpose of making them fit within 2–3 minutes, some of them, ST2 and ST3 in particular, were still slightly long (toward the higher end of the 2–3 minute range or over 3 minutes). The result of having long texts was that the entire "hometown" test was somewhat long. The length combined with the concentration demanded due to the style and content of the texts may have caused subjects to lose their concentration as the test progressed. This may have contributed to the low scores for ST3, though there is no way to be certain. It would appear that, in general, attempting to use texts that require higher levels of language ability on a RTT may have the potential to greatly increase interference in comprehension due to mental fatigue during the course of the test, thereby decreasing the validity of the test. It may have been possible to overcome any problems with fatigue, for the "hometown" testing at least, by developing additional test tapes with each tape having fewer texts. This would have increased the time and work required to develop a test substantially, however. Also, the possibility of a problem with fatigue may have been reduced by shortening the longest texts further. This also would have required additional work, and it would have resulted in less text upon which to base possible questions, which would have led to fewer questions having been developed and tested. As seen in

the previous section, there already is a problem with having too few questions for some texts, even though they were not shortened further.

It is important that a researcher doing RTT testing have a good understanding of the texts being used in order to evaluate potential questions as well as responses to questions during testing. In many cases, the researcher may not be a native speaker of the test language(s), and it is necessary to gloss and do free translations of each of the texts. Using formal, non-narrative texts made this process more difficult as I had very limited prior exposure to the vocabulary and styles used in texts ST1–3.

Questions on a RTT should only ask directly and clearly about the content of the text, they should be short and require only short answers, and they should occur on the tape very soon after the portion of text that contains the answer. It was found that, in using formal, non-narrative texts, the preparation of test questions was somewhat more difficult in each of these respects.

It is to be expected that formal texts will contain longer and more complex sentences. Many of the sentences in texts ST1–3, indeed many of the clauses, are quite long. For example, ST3.11–12 is a single (complex) sentence containing 73 words; the first independent clause in this sentence, all of ST3.11, is 35 words long. The length and complexity of sentences in these texts made it quite difficult to come up with questions that required only short answers and also to record the test tape so that each question occurs immediately after the part of the text containing the answer. It was necessary to try using some questions that related to somewhat long portions of text, though with the expectation that subjects would be able to remember only a fraction. It is not surprising that scores for some of these questions were very low. For example, for question B6 on text ST3, two subjects responded that the text was too long or that they could not remember; the overall score for this question was 30%.

Given the topics and formal style of the texts, it was more difficult to ensure that questions only ask about content directly and that they do so clearly. For example, upon reviewing results of the hometown testing, it appeared that question A11 on text ST2 did not address the content of the text specifically enough:

ST2.17 นอก จาก นี้ ยัง ดึง คน ใน ชุมชน ให้ เข้า มา มี บทบาท ใน
nɔ̂ɔg càag níi jaŋ dyŋ khon naj chumchon hâj khâw maa mii bòdbàad naj
outside from this.(AdN) yet pull person in assembly so.that enter come have role in

การ จัด การ ศึกษา ด้วย วิธี ต่าง ๆ
kaan càd kaan sỳgsǎa dûaj wíthii tàaŋ (redup)
NOM arrange NOM study also method different REDUPL

"In addition, (the Dept.) is also drawing people from local communities to take on various roles in organising education."

Q A11) ชุมชนมีส่วนร่วมในการศึกษาอย่างไร 'What part do people from local communities have in education?'

The correct answer is จัดการศึกษา /càd kaan sỳgsǎa/ 'organise education', but 4 of 10 subjects responded by naming more specific activities, e.g. "donate books", that were mentioned in the preceding units of the text. A fifth subject gave no response, and the overall score for this question was 50%.

The difficulty of developing clear, simple questions for texts ST1–3 is reflected in the rejection levels. As seen in Table 7, far more questions were rejected for the more formal texts than for the less formal texts:

14

Table 7. Question rejection levels

		# of questions	# rejected	rejected (%)
non-formal	ST/SC	16	1	6%
	ST4	28	5	18%
formal	ST1	20	9	45%
	ST2	32	11	34%
	ST3	22	16	73%

Common practice for development of a RTT is to evaluate 15–20 questions in "hometown" testing. Because of the higher rejection level for questions on formal texts, however, it must be recommended that, if possible, 30 or more questions be developed and evaluated in "hometown" testing in order to be sure of having 10 usable questions. (Note that for text ST1, 20 questions were evaluated, and there proved to be just barely enough usable questions.) Evaluating this many questions requires substantially more work in the development of the RTT, not only in the already difficult task of thinking of potential questions, but also in preparing test tapes and in testing since it would likely be necessary to produce 2 or possibly even 3 test tapes, and the number of subjects that must be tested has to be multiplied accordingly.[18]

There is an additional problem to be faced in developing questions: That the questions which are clear and simple are somehow **too** simple. For example, consider the following portion from text ST3 with the corresponding question:

ST3.3 ทาง หน่วย งาน ได้ ดำเนิน การ จัด ทำ โครงการ ศูนย์ ฝึก อบรม
thaaŋ nùaj ŋaan dâj damnəən kaan càd tham khrooŋkaan sǔun fɨ́g ʔòbrom
way unit work PAST proceed.(roy) NOM arrange make project center practice train

เกษตรกรรม แบบ ประณีต
kasèedtrakam bèɛb praniid
agriculture style precise

"Officials have made arrangements to build a comprehensive agricultural training center..."

Q B3) เขาจะทำการฝึกอบรมเกษตรกรรมแบบไหน 'What was the agricultural training program that they organised like?'

The correct answer to question B3 is แบบประณีต /bèɛb praniid/ 'precise, elaborate', and all 10 subjects responded correctly. Out of curiosity, however, I asked a couple of subjects what these words meant; they said that they didn't know. This is a problem of validity of the test: the intent of the RTT is to measure comprehension, but in this case the question fails to detect a lack of comprehension.

A further potential problem is in some ways similar to that just described: It was stated earlier that a text should not be too familiar to subjects; a subject's ability to answer each question should be dependent entirely upon comprehension at the time of the test; it should not at all be dependent upon prior familiarity with the text itself or upon familiarity with the subject matter discussed in the text. For example, a question such as *How did he cook the rice?* would be inappropriate since a subject might have been able to answer the question even without understanding the text, or even without hearing it! If simple, non-formal narratives are being used for a RTT, it is not too difficult to control this: Recounting a personal experience

---

[18] As noted in §1.3, Grimes (1993) considers it standard practice to evaluate 30 questions. This is sure to be rather more difficult for formal, non-narrative texts than for simple narratives for the various reasons already discussed.

that is not public knowledge would be sufficient for this. For formal language texts, independent knowledge about the content of a text may be more difficult to control, however, depending upon the topic of the text. One may not be able to anticipate whether a subject will be familiar with the subject matter of a text. Given the already difficult task of thinking up simple questions, one might readily forget to provide the extra degree of care required. During the current research, I was not aware before conducting "hometown" testing of any problems along this line with any of the questions that had been developed. On reflection afterwards, however, I realize that it may have been a problem is some instances. For example, ST1.5 says, *"...it is now so common... for women to engage in prostitution that it... is causing the spread of AIDS."* This is followed by the question, *"What social problems result from this type of work."* It is quite possible that most subjects could answer the question correctly even without hearing the portion of text that contains the answer.[19] There is no way, however, to be certain as to whether most subjects had answered the question based only upon what they had just heard in the text, or upon independent knowledge. The result is that some questions may be deemed acceptable though they lack validity since they fail to measure comprehension. The most that can be done to maintain validity in the test would be to reconsider each question and eliminate any (assuming that the pool of usable questions is sufficiently large) which are suspected of being answerable based upon independent knowledge. Such decisions will in many cases be purely speculative, however.

Returning to the rejection levels shown in Table 7, particular consideration must be given to the especially high rejection level for questions on text ST3: 73%, almost twice as high as the combined rejection level for texts ST1 and ST2. Given that the same difficulties were faced for all three texts, it is unclear why there was such a difference in the number of questions rejected. It is possible that the difference is coincidental and that there simply were that many more unacceptable questions for text ST3. This might be accounted for by the fact that this text is even more formal than texts ST1 and ST2 and that the problem of making up good questions was, therefore, that much more difficult. Even so, such a high rejection level is still surprising. This begs a second possible explanation: that the subjects had difficulty in comprehending and processing text ST3.

Two possible reasons could be cited for why subjects may have had difficulty in processing text ST3: The difficulties could be due to the fact that the text is in a written style but that it was being delivered orally. It is not unlikely that this did contribute to difficulties in answering some questions, but it is not clear that this alone can account for the extremely low scores and high level of rejection of questions.

The second possibility is that the difficulties are due not to the use of written style but rather to the text not being in the native language of any of the subjects. Recall that a RTT must begin with a subject control test (cf. §1.3), one of the purposes of which is to ensure that subjects are native speakers of the language assumed for test subjects.[20] If we assume, for the sake of argument, that most of the questions for text ST3 are good, then the results of the testing already done could be interpreted to mean that 16 of 20

---

[19] Text ST/SC is most likely to have had a problem of this sort during the "hometown" testing since a text was used that discussed an incident that would be quite familiar to residents of Chiang Mai, the flooding in 1994. This went completely unnoticed during the "hometown" testing, though it appears that any problem along this line with this text is, in fact, inconsequential since there does not seem to be any question as to whether or not each of the subject is acceptable as a test subject.

[20] Cf. note 14.

subjects are not acceptable.[21] Since testing with the other texts shows that these subjects do not have difficulty with the test methodology, they could only be unacceptable because they do not have native-speaker ability in the language of text ST3. Of course, we do not expect all the questions to be valid, but even after excluding the 10 questions with the lowest scores (a rejection level of 45%), 14 out of 20 subjects might still be deemed unacceptable.[22] A different way of describing this view would be to say that the "hometown" testing for text ST3 was not actually done in a "home" town!

While the data collected thus far do not establish any particular account of the extremely high level of rejection for questions on text ST3, it is certainly a reasonable possibility that this last account is a major contributing factor. This would have some very important implications: Most significantly, it suggests that for typical Central Thais, the royal range of language is effectively a second, learned language. This is consistent with the comments from Smalley 1994 cited in §1.1, that sacred and royal language are not necessarily controlled by native speakers of Standard Thai, and that they generally require considerable education in order to master them.

Interest in this research was raised in relation to questions as to whether KM-speaking Christians could adequately understand Standard Thai translations of the Bible. Since those translations make use of the royal range of Standard Thai, it was desired that the RTT being developed should, in part, specifically test the extent to which speakers of KM are able to comprehend the royal range of Standard Thai. It will not possible to address that issue at present, however, since only 6 of the questions developed for text ST3 are usable whereas 10 are required. If it is the case that most native speakers of Standard Thai are not able to understand royal language, however, then it would be expected that far fewer KM-speakers would be able to understand the royal language.

There is a problem to be faced in any attempt to test ability in royal Thai: In order to evaluate the extent to which a person can comprehend royal Thai, they must be tested using a valid comprehension test. But, in order to ensure that such a test is valid, it must first be tested on a "hometown" sample, i.e. on native speakers of royal Thai. It is not clear, though, if there are ANY native speakers of royal Thai at all. It must suffice to evaluate the test using fluent speakers (native or otherwise). There is no independent means, however, of knowing whether someone is a fluent speaker of royal Thai or not except for the subjective judgment of those who consider themselves qualified to evaluate.[23] What is especially difficult for the researcher is the fact that those who are most likely to be fluent in royal Thai would also be the least accessible. As a result, it appears that there may not be any easy way to evaluate levels of comprehension

---

[21] A threshold score of 85% was used for this calculation. For a threshold of 70%, 12 subjects would still be deemed unacceptable.

[22] With a threshold of 70%, only 4 subjects would still be deemed unacceptable. It appears that a 70% threshold is too low, however, since if these 4 subjects had been eliminated for the "hometown" testing on text ST3, it would still remain to account for an extremely high question rejection level: 17 out of 22 (77.3%) questions rejected. Clearly, these 4 subjects alone do not at all account for the low scores for the questions on this text.

Raising the threshold above 75% would result in 8 subjects being deemed unacceptable. These would consist of most of the youngest subjects and most of the subjects with the lowest education, which would be consistent with a claim that the language of text ST3 is learned rather than being the mother tongue of the field of subjects tested. This discussion, however, is only conjectural as valid data has not been compiled in support of the suggestions being made.

[23] In the usual case with a more typical variety of speech, the tacit basis for assuming fluency is that there is a sizeable body of people who are known (or can reasonably be assumed) to be native speakers. This is not possible for royal Thai.

of royal Thai for any but the smallest of samples; at the least, it would be rather difficult to develop a RTT that includes texts that use royal Thai.

The question as to the extent to which speakers of KM (or even native speakers of Standard Thai, for that matter) are able to understand texts that use royal Standard Thai will, therefore, remain unanswered at this time. In spite of this uncertainty, royal Thai is predictably used is particular contexts, either with the assumption that the average member of the target audience is adequately able to understand, or with the attitude that the use of royal language is appropriate in those contexts regardless of whether the target audience can understand or not. If a value is to be placed upon communication of some particular message to the target audience, it appears likely that the use of royal language would conflict with that goal.

4. Conclusions

The process of developing a RTT for testing bilingualism in Standard Thai among speakers of KM has been documented, and a sufficient number of usable questions for texts ST1, ST2 and ST4 have been identified. Text ST3 cannot be used in bilingualism testing at this time, however, since too few of the questions for this text proved to be usable. Various difficulties in developing a RTT using formal, non-narrative texts have been identified.

Some of the problems encountered in developing a RTT using specifically formal, non-narrative texts were of a sort that could render some of the questions used on the test invalid; i.e. some problems may result in questions not measuring level of comprehension, which is purpose of the test. It was mentioned in §1.2 that some have suggested that a RTT may be made more suitable for bilingualism testing by the use of texts which require greater levels of proficiency, but also that no studies in which this was actually attempted are known of. Difficulties encountered in the current research indicate, at best, that attempting to use the RTT methodology in this way will be somewhat problematic. At worst, they suggest that using the RTT methodology in this way might not be entirely valid. There are other issues that are relevant to questions regarding the validity of the use of RTTs in the way described here; these are beyond the scope of this paper.

The small number of questions developed for text ST3 which proved to be usable has led to the suggestion that the royal range of Standard Thai may not be anybody's mother tongue and was not fluently understood by the native speakers of Standard Thai used for "hometown" testing. This led further to observations regarding problems that would be anticipated with any attempt to develop a RTT that measures level of comprehension of royal Thai.