2014

# The ABD of orthography testing: Practical guidelines

Elke Karan
*SIL-UND*

Follow this and additional works at: https://commons.und.edu/sil-work-papers

Part of the Linguistics Commons

## Recommended Citation

# The ABD of orthography testing: Practical guidelines

Elke Karan

Literacy and Education consultant, SIL International

*Developing an orthography for a previously unwritten language may prove to be more of a challenge than anticipated. Orthography testing can help field workers and the speech community identify issues which warrant special attention. Subsequent testing will reveal whether or not various factors have been taken into consideration and challenges overcome. Findings might indicate that the orthography functions well or that it requires additional thought, revision, and /or additional testing.*

*Few how-to helps on orthography testing exist. Drawing from various resources, the author provides practical guidelines and resources for field workers who are planning to engage in orthography testing activities.*

## 1.  Introduction

Orthography design, a vital component of language development, is rarely a straightforward process. A variety of factors come into play and might complicate efforts. These may include national policies, linguistic complexities inherent in the language, complexities inherent in the script that is to be adapted, precedents set by other languages which use the script, a competing system already in use by some of the stakeholders, aesthetics (acceptability of the look of the system), identity issues that shape expectations and influence acceptability, a complex dialect situation, and technical considerations such as ease of texting, for instance. (See Cahill & Karan 2008.)

The idea that newly developed orthographies ought to be tested before they are implemented makes perfect sense. It would be valuable, as well as satisfying, to have evidence that the various issues have been adequately dealt with, i.e., that the orthography under consideration is easy to learn and use and that it meets with the speech community's approval. Ideally, such supporting evidence would be assembled *before* a formal proposal is submitted to the appropriate authorities for final approval. Powlison warns, "The longer an orthography has been in use, the more sacred its form is held by its users. Consequently, if changes are anticipated, they should be undertaken as early as possible in the life of the orthography. It is ideal, though not always possible, to have the orthography tested and proven satisfactory before any books are printed" (1968:85).

Few how-to helps on orthography testing exist. Yet it is likely that orthography testing is more commonplace than the small number of published articles on this subject would suggest. A problematic issue may simply trigger a knee-jerk reaction: "This doesn't work well; let's try something else." The issues and subsequent attempted solutions may not be considered noteworthy, thus the testing efforts, findings, and rationale behind decisions may go undocumented and thus escape notice by those who might benefit from such information. For example, Curtis Cook's orthography testing effort among the Zuni [zun] in New Mexico would have gone unnoticed had others not taken note and written about it (Walker 1969; Venezky 1970; Berry 1977).

Those who have carried out orthography testing and written about it in the past have done others who follow a great service. Lessons can be learned from their investigative questions, observations, findings and procedures. Those planning to conduct orthography testing would do well to review and learn from what others have done and written and to document and publish their own experience.

My earlier article on orthography testing (Karan 2013) advocates for orthography testing, includes a literature review, and addresses the topics *Why test, What to test,* and *Who to test*. This current article

more closely examines the practical question of *How to test*, thus complementing the first. Although some qualitative approaches are mentioned, the main focus of this paper is on quantitative research.

Section 2 of this paper presents orthography testing as an evaluative process which is to be considered an integral part of a language development program.

Section 3 discusses common investigative questions related to orthography testing.

Section 4 describes a variety of testing activities which can be carried out throughout the orthography development process.

Section 5 gives detailed instructions for preparing for and carrying out orthography testing at scheduled events. It also addresses the question of how many respondents to test.

Section 6 highlights characteristics which distinguish formal testing from informal testing.

## 2. Planning for orthography testing

### 2.1 Orthography testing as evaluation

Orthography testing is a type of evaluation. Evaluation is a necessary component of any project, program, or service. Stakeholders, i.e., those with a vested interest, need to know if goals are being met and services are accomplishing what they were designed to do. In education, for instance, evaluation provides evidence that helps determine if teaching and learning are on track. Evaluation is not just about identifying deficiencies, but also strengths. Its aim is to discover what's working, what's not, and why, and to work out suitable remedies.

Evaluations don't just happen. They need to be planned and treated as an integral part of programs and processes. They may be seen in a negative light, i.e., as too costly in terms of time, energy, and funds required. However, the benefits of evaluations, including orthography testing, far outweigh the cost.

Evaluations if done well[1]

- add understanding
- provide accountability
- identify strengths and evidence of effectiveness
- identify ineffective practices, paving the way for improvement
- provide needed documentation
- enhance credibility.

But evaluations lack substance if the most basic components, namely, clear aims and objectives, are missing in planning. If objectives and corresponding indicators, i.e., appropriate observable and measurable evidence, were never identified, there is no basis for evaluation. So, early in the language development process, workers need to draw up specific objectives for the orthography design process and for the orthography itself.

Some objectives and related questions generate more tangible evidence than others. For instance, attitudes toward the use of diacritics or whether or not certain symbols cause people to stumble while reading are probably easier to ascertain than "how partnerships are doing" or "how community ownership of the process is coming along." Yet, it is possible to evaluate less tangible desired outcomes as well, as long as they were enumerated among the objectives. If objectives were formulated, indicators can be drawn up to help determine later whether or not the objectives were met.

Once indicators have been defined, appropriate assessment instruments can be developed. Testing can thus be carried out, followed by analysis of the findings and then reporting. Consequently, if deficiencies in the orthography are discovered, discussions of potential corrective strategies would ensue and

---

[1] This list is based on one drawn up by the Pell Institute and Pathways to College Network (2014).

adjustments would be made. If the objectives remain stable, the process can be visualized as follows:
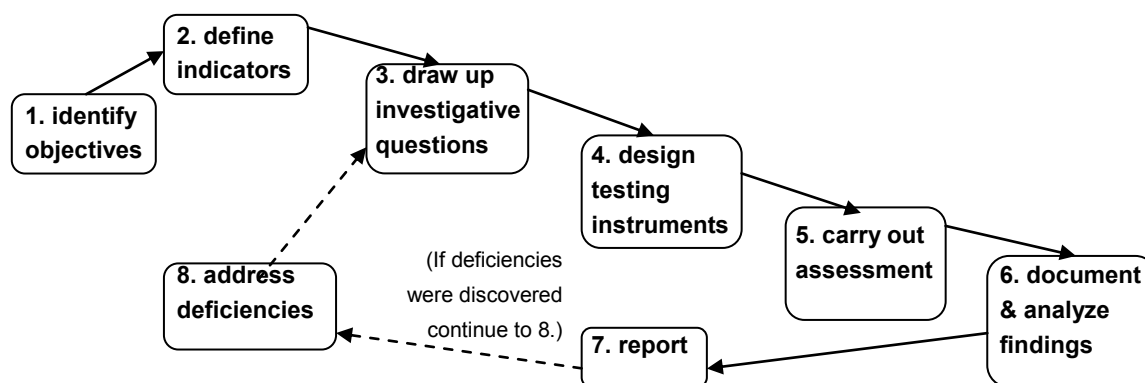


**Figure 1: Model of the orthography testing process**

*2.2    Objectives and indicators for newly developed orthographies*

A variety of authors have discussed criteria for optimal orthographies over the years. A useful list of qualifications developed by William Smalley (1959), commonly referred to as *Smalley's Maxims,* continues to be cited and discussed more than fifty years later. (See Hinton 2014, for example.) Objectives for orthography development, rather than being language-specific, are usually quite similar from one language to the next:

- Linguistically, the orthography should represent the language well, with attention given not only to the sound system, but also to morphology and syntax.

- Sociolinguistically, the orthography should correspond with mother tongue speakers' desired identity/identities and extend to different speech varieties, where feasible and desired.

- Politically, it should gain approbation of the authorities and serve as a unifying factor, and hopefully also find favor across the speech community. Attaining consensus would be ideal, but this ideal may be elusive in the early stages of development and implementation. Standardization may take some time.[2]

- Educationally, the orthography should be easy to teach and learn. This includes reading as well as writing. It applies to the formal realm of schooling, as well as the non-formal realm, i.e., with youth and adults. Non-literates should be able to learn and use the system without struggling. In addition, for those already literate in a given language, it is desirable that their literacy skills would transfer easily to this additional language. This is most easily attained if the script is the same and if there is harmonization between the systems minimizing interference between them. Ideally, teachers will readily learn and embrace the system, especially if the language will be taught as a subject or used as a language of instruction.

- Aesthetically, the orthography needs to appeal to intended users. It should not have anything that would be stigmatized, make it appear ugly or difficult in the eyes of prospective users.

Ideally, these factors should contribute to enhance motivation to learn and use the system. *Maximum Motivation* is the criterion that heads Smalley's list.

What types of indicators should be chosen to determine if these desired objectives have been attained? The kind which are *observable* in some way. Approval rates and lists of objections can serve as indicators to measure the degree of acceptance. Errors and disfluencies in oral reading, such as hesitations

---

[2] For a discussion of the standardization process for newly written languages, see Karan 2014. Standardization: What's the Hurry?

and repetitions, the time taken for literacy tasks, correct or incorrect answers to comprehension questions, and consistency vs. inconsistency in spelling can serve as indicators for learnability and efficiency of the system.

Those evaluating an orthography must beware of indicators which might actually be inappropriate in an evaluation. For instance, someone might suggest *literature sales* as an indicator of orthography acceptance and motivation to learn and use it. But literature sales, although easily quantified, are not necessarily a reliable indicator, because other factors, not related to the orthography, may be at play. For instance, potential clients may not have discretionary funds to spend on non-essentials; or the content of the literature produced may not interest them; or literacy rates may be low, limiting the prospective client pool. Even the color chosen for covers could positively or negatively impact sales.

It may also be possible that underlying concerns impact the use of an orthography. For instance, if educators do not embrace the orthography or if they claim that it is difficult, they may in fact be resisting the local language's integration into the school curriculum, and not the orthography itself.

I personally encountered such confounding factors while involved in the Sango literacy project in the Central African Republic (CAR) in the 1990s. I wrongly attributed the lack of literature sales to resistance to the government-imposed 1984 orthography.[3] The actual problem, however, was financial: low cost materials, such as news sheets and paper brochures actually sold well. Trying to keep cost down on literature, the project primarily produced A6-sized booklets, since this cut the need for costly cover stock in half. This actually worked against sales. Surveying a variety of individuals about their attitude toward the orthography, I accidentally discovered that some of them did not consider A6-sized "books" *real* books. In their mind, the minimum size to qualify as a book worth owning was A5.[4]

## 2.3    Timeline, funding and expertise

Orthography testing is research. Research requires planning, effort, and funding. Since orthography testing will not happen unless proactive plans are made, field workers should explicitly integrate the process and testing events into the yearly and longer range plan, i.e., three-year, five-year, or whatever time frame is customary. This will enable allotting appropriate time periods for necessary activities, events or travel and budgeting for them, i.e., allocating funds for associated expenditures.

Those intending to do orthography testing need to know their strengths as well as their limitations. Orthography testing relates to linguistics, sociolinguistics and reading theory. Specialists from these domains should have input into the planning and research design. Especially for formal testing, someone experienced in doing research in the social sciences and in statistical analysis should be recruited to serve on the team. If this is not possible, a consultant with such expertise could participate virtually, providing input into the preparatory process and later into the analysis and interpretation of the findings. It is not uncommon to see three or more authors' names attached to quality research reports.

## 2.4    Evaluation approaches

Most evaluations take one of two philosophical approaches: a summative evaluation, or a formative, integrated, ongoing evaluation. The first involves a scheduled event toward the end of a program cycle and is designed to discover if something worked or if the program delivered what it promised. The second is a continuous monitoring process, designed to provide data along the way that will indicate what successive improvements should be made. Effective evaluations usually use both

---

[3] Sango [sag] is the official national language of the CAR, alongside of French. For a discussion of the Sango orthography reform and description of informal orthography testing efforts carried out from 1991–2001, see Karan 2006, chapter 12.

[4] A5 is half the size of an A4 sheet, while A6 is a quarter of the size of A4. The dimensions of A6 are 10.5cm x 14.8cm.

approaches. For instance, the David and Lucile Packard Foundation (2014) believes in the necessity of summative evaluation as it relates to accountability but also promotes "real-time" monitoring, "encouraging a culture that 'thinks evaluatively' throughout … planning, implementation, monitoring, assessment, and course correction."

So it should be with orthography development. Thinking "evaluatively" throughout the process will guide decision making. Early on, formative testing will identify what needs attention and provide the baseline data for later comparison. Integrated monitoring will help in decision making and lead to improvements. Later testing events will answer the question "Are we there yet?" i.e., have we met our objectives? However, with orthography testing there is no absolute final summative evaluation. Languages change. Social conditions, identities and motivations change. Therefore, revising an orthography should always be an option, even if it has attained standardized status.

### 2.5    *Informal testing vs. formal testing*

Formal orthography testing resulting in published reports which inform the academic community is to be encouraged. But for pragmatic reasons and simplicity, informal testing has tended to be the default approach for orthography testing. It is adequate in most situations and will provide answers to common orthography-related investigative questions. During the initial planning, the field worker must take into consideration that informal research is generally not viewed as scientific. It is considered subjective since it is usually not based on random sampling and systematic data gathering and because the interpretation of findings is often based on intuition rather than statistical methods. Informal research usually consists of activities involving observing, counting, and/or discovering opinions, which are not usually considered reliable research methodologies, and thus findings are not considered replicable.

Yet, informal testing can, in fact, be quite rigorous and uncover valuable information. It may be all that is needed in certain situations. However, in others, informal testing ought to only be considered the starting point. It will likely reveal additional research questions that may call for formal research.

The complexity of the research question alone does not determine if informal or formal testing is more appropriate. Rather, additional deciding factors are a) who needs and wants to know and b) what it will take to convince them. Formal testing is the better option when it is necessary to "satisfy a critical audience or to help resolve an issue over which there is much disagreement or controversy" (Karan 2013:11). (See section 6 for a brief discussion of characteristics specific to formal testing.)

Taking an informal approach for some of the research questions and dealing with more controversial questions via a formal approach is perfectly acceptable. In this article, the reader should assume that the discussion applies to both informal and formal testing, unless specific reference is made to one or the other.

## 3.    **Investigative Questions**

Investigative questions have to be clear in the researcher's mind, since they will guide the research strategy. For instance, they will affect activity planning, the design of test-instruments (research tools such as questionnaires, the choice of reading and writing tasks, etc.), sampling choice and size, permissions needed, and analysis of the findings.

The research question "Do the majority of X speakers prefer to write /u/ as <u> or as <oo>?"[5] clearly needs a different research design than the questions "Can mother tongue speakers of X who are fluent readers of English learn to write the proposed X orthography with four hours of instruction? If not, what symbolization issues are problematic for them?"

---

[5] In this paper, oblique slashes / / denote contrastive sounds of a language, while angled brackets < > denote orthographic representation. Capital letter *C* denotes any consonant in the language.

Certain investigative questions will need to be raised in just about every orthography development situation, but they need to be fine-tuned according to

- the language structure and its specific complexities
- the scripts being considered
- the language(s) of wider communication in the area
- the language varieties occurring and the degree of difference between them.

Other factors that might affect investigative questions as well as research design, might be the literacy rate in the community or whether the language is highly endangered and has few speakers left.

Karan (2013) briefly discusses ten orthography issues which commonly need to be tested. They are listed here, coupled with potential investigative questions.

- *Script preference:* Do the majority of ___ speakers prefer the ____ script or ____ script for writing their language?
- *Symbolization of certain contrastive sounds:* Do the majority of ___ speakers prefer to write /___/ with <option 1> or <option 2> or <option 3>? Do options 1, 2, and 3 work equally well for reading? For writing?
- *Proposed underrepresentation of phonemes:* Will underrepresentation of vowels /__, __/ allow the orthography to be extended to all the dialects? Will it pose a problem for some?
- *Diphthongs and semi-vowels:* What do people prefer: writing labialization with <Cw> or <Cu>? Is reading performance affected by one symbolization as compared to the other?
- *Cross-syllable graphemic ambiguities:* Which symbolizations cause readers to parse words incorrectly?
- *Stress:* What is the functional load of stress? Will marking stress result in fewer reading errors?
- *Tone marking options:* Which tone marking notation is preferred by speakers of ___. Which tone marking system results in the least amount of errors and the best fluency rates?
- *Word breaks and the use of white space:* How do speakers of the ___ language feel about inserting extra space between words since this is not usually done with the ____ script?
- *Elision and unusual consonant sequences:* What proportion of ____ speakers prefer to write the elided consonant ____, as compared to those who prefer to not write it?
- *How to write loan words:* Do speakers of ____ wish to preserve the spelling patterns of International Language loan words in ___ texts or do they wish to adapt them?

Some of the issues will have multiple investigative questions associated with them. For instance, for script preference, the primary investigative question might be formulated as follows: "What script do the majority of adults prefer for language X?" If there is a divergence of opinion, another might be, "If reading were taught in language X in primary schools, in which script do parents prefer their children to learn to read language X?" Answers to these questions can be discovered through opinion polls, surveys, or interviews. For the second question it may be helpful to show respondents sample stories suitable for children in the various scripts under consideration.

Regarding underrepresentation, one investigative question might be, "Do the majority of mother tongue speakers want sound 1 to be written with a separate grapheme from sound 2 or not?" If they do not, an accompanying question might be, "What grapheme do they prefer for writing the two sounds?" Since underrepresentation is not ideal, it is wise to follow up with performance and comprehension tests. The investigative follow-up question would be, "Does the underrepresentation of sound 1 result in ambiguities that interfere with reading fluency and comprehension of texts?"

We note that investigative questions often relate to these key areas:

1. people's general preferences, often based on identity issues and expectations and usually influenced by other writing systems in use in the area;

2.      people's specific symbolization preferences;
3.      comparative efficiencies of symbolization options;
4.      the efficiency of the total system, i.e., the effect of the sum of individual orthography decisions.

Issues of a more complex nature may also need to be tested. An example of such an issue is whether to write the output of phonological rules, i.e., the surface spoken form, or to write the underlying form. (See Snider 2001; 2014a. See also Roberts, Snider & Walter (forthcoming)). Mother tongue speaker perceptions and their awareness of some phonological processes vs. others are not always easily ascertained. It is therefore necessary to discover mother tongue speaker intuitions. Observing actual writing and oral reading and listening to mother tongue speaker objections to certain suggested symbolization options may reveal them.

## 4.  Testing activities

A variety of testing activities are commonly used to find answers to investigative questions. Moe (2005) offers some practical suggestions for testing during various stages of the orthography development process. Although the focus is on testing Bantu language orthographies, the activities listed are applicable more broadly. For instance, field workers are encouraged

- to collect and analyze existing print materials to look for inconsistencies in writing
- to seek out individuals who have some experience writing the language and ask them what problems they have encountered and what puzzles them
- to organize writers workshops and look for systematic errors writers make that might indicate an erroneous linguistic analysis
- to test people's ability to read the orthography in order to discover any deficiencies
- to test people's reading of different orthographic options to discover which work better
- to evaluate mother tongue speaker reactions to the orthography
- to test the orthography in trial literacy programs.

Some of the activities listed will be discussed in more detail. A combination of activities can be used to strengthen findings for or against certain symbolization options.

Some types of activities provide anecdotal data, i.e., findings based on stories or testimonies which carry weight but are not quantifiable. Others provide qualitative data, i.e., that which can be observed but not measured. Some result in quantitative data.

Qualitative data has often been considered inferior to quantitative data. Mark Karan (2009), drawing from Berkowitz (1997) and Trochim (2006), points out that both qualitative and quantitative research can be used for various research questions. The two approaches are to be considered complementary. The author clarifies that the opposite of good quantitative data is not qualitative data; the opposite of good quantitative data is bad quantitative data. The opposite of good qualitative data is bad qualitative data. Karan (ibid.) states that bad data results not from the research approach itself, but from a researcher's incorrect assumptions, poor sampling and poor measurements which skew the results. It is therefore important to minimize these types of risks. The main strength of qualitative data is that it provides insight into people's behaviors and perceptions. In orthography testing this relates most closely to preference and attitude testing. Karan lists commonly used qualitative methods: observation, interviews, focus groups (group interviews), and studies of existing documents and/or recordings. We note an overlap with Moe's suggestions. A more detailed discussion of some of these activities follows.

*4.1    Observation and analysis of writing samples*

If an orthography is already in use, observation of native speakers' reactions to it and an analysis of writing samples can provide a surprising amount of information. Letters, personal notes, radio or TV scripts retrieved from waste baskets, public signage, etc., indicate actual practice and reveal areas of difficulty. Unmonitored writing by "naïve" individuals, that is, people writing freely without someone compelling them to follow certain conventions, will give insight of a different kind than writing done by those who have been trained in the orthography. For example, submissions to a story writing contest may reveal more about writers' intuitions and preferences than texts written at a workshop where "how to write the language" is one of the instructional modules. This latter situation might reveal more about learnability and ease of use of the system. However, if the orthography as presented is not being consistently applied, it might also reveal how deeply ingrained orthographic rules are from a language of wider communication or language of education or point to a faulty linguistic analysis. Observation in itself does not explain such root causes. It does, however, point to additional research questions that need to be asked.

A variety of training events present opportunities for observation. Personal information about the individuals whose reading and/or writing will be analyzed can be collected in advance, through the registration process. Events which offer a chance to observe include 1) skill-transfer workshops during which individuals, literate in a language of wider communication, are introduced to the preliminary orthography of the language under development, and learn to read and write it; 2) teacher training sessions for formal or non-formal programs; and 3) writer training workshops. Reading and writing tasks—composition as well as dictation—should be included among the activities.

Some participants will struggle more than others. The workshops will provide opportunities to analyze the root of difficulties and to invite popular participation in the solution-finding process as various options are discussed and tried. Maybe a change is not what's needed, but rather a better teaching strategy or a special mnemonic device.

In my years of experience with the officially decreed Sango orthography and its implementation, two particular opportunities to observe stand out: At a writers workshop, two linguists from the national university and a choir director were able to mark tone relatively consistently, while the other fourteen participants, although making an effort, found marking tone on even the most common words a great struggle. Comparing tone levels syllable by syllable against standard one-syllable word models—a very tedious process—yielded poor results.[6]

An unexpected delivery provided the second opportunity: A leader in the literacy project surprised me with a piece of paper on which announcements from a Sango news program were written. He had visited a friend at the TV station and had rescued it from landing in the trash. Those working in mass communication, i.e., radio and television staff, had received the most rigorous training available in the official orthography. Yet, here was a piece of paper providing evidence that tone was *not* being written by the "elite," even though they were urged to and had been trained to do so.

Since a person can only be in one place at a time, it is advisable to train several individuals to gather helpful information. The burden should not fall on one or two researchers. It would be beneficial to provide literacy teachers and trainers of trainers with small notebooks in which they would note community member attitudes and observations of common learner hesitations or errors.

---

[6] The awareness that native speakers have of tone patterns as opposed to individual tones, determines how to best represent and teach tone in a tone language orthography. (See Snider 2014b, chapter 5 and Roberts 2011.)

*4.2    Interviews and questionnaires*

Interviews allow researchers to build relationships with stakeholders. They are effective for gathering information and gaining insight. They are usually conducted face-to-face and one-on-one. This is positive: the process makes individuals feel validated since their input is valued. Interviews can be quite casual and impromptu or, at the other end of the spectrum, quite formal, necessitating an appointment. This may depend on the status of the interviewee and the working relationship between them and the one designated to do the interviewing.

Well-conducted interviews require an adequate trust level on the part of the interviewee and ethical consideration on the part of the interviewer. They can reveal emotive issues, such as preferences, prejudices, frustrations with certain symbolizations, or reactions to the decision making process itself. Such information is helpful and should be noted since affective filters can work against the acceptability or learnability of a system. Smalley (1964:14), well aware of such issues, comments, "The ways in which these emotional factors will enter into the acceptance of a writing system in areas where new systems are being devised, remains one of the great critical problems in this field. Some workers have paid altogether too little attention to the problem, with occasional lamentable results."

Although emotional reactions must not be ignored, it is important to note that they do not negate the potential efficiency of a system. It is therefore wise to prepare questions ahead of time about what specific symbolizations are acceptable and which ones pose a problem. This is especially important in an orthography reform situation since change is never introduced easily. Highlighting what is actually a sore point in the proposed system may help prevent the proverbial throwing out of the baby with the bath water. It is appropriate to ask, "What do you like about this way of writing?" It is not necessary to only focus on the negatives.

Nor should one assume that a negative reaction is deeply rooted. It may be a prejudice stemming from lack of experience or understanding. I once interviewed an individual who was accustomed to reading Sango in an older writing tradition and showed her a sample text using the revised orthography. "It looks like Lingala," she commented.[7] Her intonation revealed that she had an initial negative reaction to the new system. But when I modeled reading a couple of sentences and encouraged her to read on from there, her initial attitude changed, and she commented, "Oh, it's easy to read. That's good."

Interviewers should not ask *leading* questions which influence the answer, i.e., ***not*** "Do you think people will have trouble switching to writing just <u> instead of writing <ou>, like French does?" or worse yet: "Don't you think that…?" Rather, the interviewer might say, "Look at this sample text. I'll read it out loud. Can you point out what words or letters might cause a problem for people who learned to read in the old system?" or, "Would you please read this and tell me your reaction?"

Interviews do have weaknesses. For one, they are time consuming. Information is needed from a large sample, so additional approaches are needed. For instance, interviews could be conducted with a group, with individuals noting their responses on a questionnaire. Although this approach increases the number of respondents that can participate, it has its limitations: 1) only individuals with more advanced literacy skills can participate; 2) the choice of the language or dialect used for the questionnaire might limit who can effectively participate; 3) relationship-building with respondents is less likely to occur at such an event; 4) group discussion may ensue, influencing responses; 5) questionnaires do not necessarily invite frankness; 6) forms often have blanks that only allow brief answers or else provide predetermined responses with a limited number of options to choose from; 7) even if respondents are invited to contribute longer answers, or additional options, their writing skills may not be equal to the task.

The limitations of questionnaires should not discourage their use. But clearly, their design requires much thought and care. This relates not only to asking good questions and avoiding loaded and leading

---

[7] Lingala [lin] is a language of wider communication spoken by over two million people in the Democratic Republic of Congo. (Lewis, Simons, and Fennig (eds.). 2014)

questions, but also to gathering pertinent information on respondents that will help identify trends in their responses. Illustrative words, sentences, or texts that are included need to be chosen carefully and presented error-free. (See Appendix A for a sample background questionnaire.)

Some types of questions must be avoided in test design; text books on quantitative research warn against using them. They provide data that do not easily lend themselves to statistical analysis since they are not in a comparable numerical format. Mark Karan (2009), for instance, discusses questionnaires that are contraindicated. They use continuum data in which the scale is not equidistant between intervals and the numerical format is therefore not comparable. Consider this option series:
1. always  2. frequently  3. regularly  4. sometimes  5. occasionally,  6. rarely  7. never
We note that terms like "regularly" and "frequently" are open to personal interpretation. They should be specified further: for example, *every day*, rather than *always*; *4-5 times a week*, rather than *frequently; 1-3 times a week,* rather than *regularly*, etc. Still, answers cannot be analyzed in terms of the numbers 1–7. They are only label identifiers, associated with the various options. In addition, there is no equidistant interval associated with the categories listed, making a numerical comparison inappropriate. However, counting how many respondents chose which answer is not inappropriate.

### 4.3    Feedback from users

Those who are to use the orthography are major stakeholders. Whether they are consumers or producers of materials, casual users, or teachers of the system, they will have opportunities to make useful observations regarding their own experience or that of others. Whether users are people of high status, such as language committee members, or average individuals who accidentally come across a brochure and try to read it, they need to be given a voice if they feel they have something to say.

Decision makers or mediators should therefore offer an opportunity to give feedback to those who are being introduced to the orthography for the first time as well as to those who use it quite regularly. The speech community needs to know that the orthography is to serve as a tool, that it is not yet fixed but needs to be evaluated and improved upon if necessary.

Solicitation for feedback can be presented at workshops or training events. Special announcements can be made on the radio or television or posted on websites. An invitation for feedback with contact information can be included as front or back matter or as an insert in pedagogical and other print material. This would be particularly appropriate in a self-instruction orthography guide. (See Appendix B for sample requests for feedback.)

As already noted, feedback from teachers is invaluable. Since they are usually literate in a language of wider communication, their personal reactions are likely to focus on transfer issues. Their observations in the classroom, however, are more likely to focus on learnability issues.

To be of maximum benefit, feedback should be accompanied by basic background information concerning respondents or the person observed since this may shed additional light when analyzing their input.

### 4.4    Surveys

Survey, as a research methodology, can be used for informal or formal testing. It usually suggests a large, broad sample of respondents. It presumes that certain demographic information will be collected and used in the analysis. Written orthography surveys are commonly used to determine script and symbolization preferences.

Procedures need special care. It is important to assure survey respondents that their names and responses will be kept confidential, allowing them to respond honestly. In fact, each person and his or her data should be assigned a numerical or other type of code.

Surveys tend to be less intimidating for participants if done in a group setting. The advantage for the researcher is that this approach is more economical than interviewing, since several individuals can be

"tested" at the same time. But, this also leaves room for sloppiness: Not everyone may have understood the instructions well; survey takers might skip some of the pertinent background information questions; eager participants might rush through the survey not giving attention to detail.

To avoid such traps, it is best to walk the group through the survey step by step, making sure everyone understands what is expected and takes time to think about the various options. When doing this, it is important for the surveyor to not let voice inflection indicate his or her preferred choice among the options. Spacing between survey participants should be such that copying would not be possible. A survey is not designed for working toward consensus via discussion; individuals should focus on their own papers and express their own opinions.

A preference survey has to be designed with care. The most important rule is to test for one issue at a time. No test item should try to cover two investigative questions. If at all possible, when testing one issue, sample words containing another potentially problematic symbolization should be avoided. For example, if there is a need to test how to best symbolize /ʃ/ as well as how to best represent nasalization on vowels, words chosen to test /ʃ/ should not contain nasal vowels. Conversely, words chosen to test nasalization symbol choices should not also contain /ʃ/.

Respondents may not have experience with a variety of test approaches. Underlining or circling a preference is the most straight-forward response and easiest for participants. Requiring participants to indicate their preference by filling in a blank space in a sentence in reference to a list of words variously spelled is a much more abstract task. It is not likely that there will be adequate advantage in adding complexity to the procedures of a survey.

A survey could potentially be designed in a way where various illustrations are presented one at a time and respondents are asked to indicate the spelling options they prefer for each item.  However, the use of illustrations presents a risk: recognizing pictures may be difficult for semi-literate individuals because two dimensional representations are often quite abstract and can be misleading. If illustrations are used, they should be identified by the surveyor to avoid problems.

Results will be more reliable if the various symbolization options do not always appear in the same order or in the same column. (See examples in Appendix C.)

*4.5    Analysis of the findings*

Surveys yield quantitative data. If a certain number of survey respondents prefer one option over the others, a test for statistical significance should be applied to see if the findings could possibly be the result of chance. Correlations between preferences and background information can also be tested for statistical significance.

Some data from qualitative research is quantifiable. Interview data and responses to questionnaires can be sorted and counted and can thus be analyzed numerically. For instance, pages and pages of prose, when quantified may reveal that 79% of the respondents expressed a dislike for the use of diacritics above vowels. This may lead to a decision to avoid the use of diacritics.

Testing for statistical significance concerns quantitative data as well as that which has been coded quantitatively, findings from informal testing as well as formal testing. (See sections 5.7, 6.2, and 6.3.)

## 5.   Performance testing events

Performance tests are frequently used
- to identify deficiencies in a tentative orthography
- to compare the efficiency of various symbolization options
- to evaluate an orthography's learnability and efficacy as a whole.

An orthography should allow users to perform well in both reading and independent writing. It is therefore important to include reading as well as writing activities when doing performance testing.

Informal performance tests can be integrated into scheduled events such as teacher training or writer training workshops. Much can be observed and learned even though there may not be much background diversity among attendees. To test a more diverse group and assure a broader spectrum of people in the sample, special events consecrated to orthography testing need to be organized. For instance, the Sô [sss] orthography committee in Thailand decided to first test the orthography within their small group to discover and address problematic issues. Once they themselves were satisfied, they envisioned doing testing in the larger community (Markowski 2009).

Informal performance testing can be done at the word, phrase, sentence, and/or text level. Formal testing usually requires reading at the text level and writing at the word and sentence level. The investigative question will determine what is an appropriate test design. Since performance testing activities are potentially more stressful than interviews or surveys and may bring to mind negative school experiences, a special effort should be made to put participants at ease.

## 5.1 Advance preparation of tools

After objectives, indicators, and investigative questions have been specified, appropriate testing instruments need to be devised. They must be evaluated before they are used for actual orthography testing. By way of some limited trial runs, it needs to be determined 1) if they will provide the data that is needed to answer the investigative question(s); 2) if they are easy to use; 3) if they could be shortened or simplified to avoid fatiguing respondents without sacrificing valuable data; and 4) if they contain any errors.

Typographical errors are particularly regrettable if found in texts that are intended for a timed reading, in which the number of reader errors serve as indicators of orthography efficiency. Trial runs and/or a person with an eye for detail and might catch and eliminate such flaws in testing instruments, benefitting test subjects and those doing the analysis of the data later.

Special care needs to be taken for designing tools for testing the orthography with subjects just acquiring literacy skills. Determining if an orthography is easy to learn requires that some testing be done in lower primary schools or adult literacy classes. Conducting a valid test in such settings may be more complicated than one might think. Other factors may come into play. For instance, the instructional methodology used may facilitate only global recognition of words and the memorization of texts rather than teaching word attack skills, or teachers may vary in skills and/or motivation and attitude toward the language in question affecting learning that takes place. Such factors may determine whether or not a set of learners would be a suitable test group. If yes, test instrument designers need to make sure that the orthography will be tested, and not the learners' skill level. Activities must not be at frustration level or even at the learner's recently attained skill level. Instead, they should target a level that has been mastered so it will be more likely that errors, hesitations and restarts are likely due to orthography issues.

## 5.2 Training of staff

Depending on the component activities, orthography testing can be a big undertaking. It is usually necessary to train staff members for the various tasks. Cultural learning styles, as well as preferred personal learning styles that may be present among staff, should be taken into consideration for the training. Oral explanations may be adequate for some, but it is much safer to take time to demonstrate and practice the procedures. In most cases, those carrying out the testing have never done anything that requires this degree of exactitude.

Orthography testing usually requires a certain amount of multitasking. This calls for confidence and calmness under pressure and in the face of unforeseen circumstances. Using equipment such as recorders and a stop watch may be new to an assistant's experience. Test organizers would do well to have staff practice all activities on one another and others who are not expected to participate in the real testing activities. It is wise to separate the role of test giver, i.e., the person who relates to test takers, and the role of technician(s), i.e., those who handle the stopwatch and recording device so one person does not end up

with a near-impossible task of handling everything. In preparation for testing, it is of course important to verify that recording equipment is in good working order and that adequate supplies, including batteries, are on hand. David Roberts (p.c.) believes in preparing for contingencies by training two individuals for each task and having two working recording devices ready for use. Depending on what type of recording device is used, a stopwatch may not be necessary.

Orthography testing may involve relating to a wide range of people. It will expedite the process if staff members have good people skills, showing respect both for those in powerful positions as well as for those who have been disadvantaged in life. It is important that staff workers appreciate the participants' willingness to serve in this way and not take involvement for granted. They may need to be made aware that the experience may be unsettling for test subjects. Public relations experience, a good sense of humor, an encouraging attitude, and a calm disposition can help foster relationships and facilitate the testing process.

Official paper work may be required such as a travel order for test givers coming from a distance, letters of authorization to present to local and regional officials, and, possibly, permission to hold a public gathering. If school children are to be tested, parental and school official permission may be required. If school facilities are to be used, authorization from the Ministry of Education may also be required. Participation by an agency's official representative may be expected. All this will need to be researched and arranged well in advance. Expatriate researchers would do well to seek advice from their local peers on protocol issues.

*5.3    Preliminary activities:*

It is important to make people feel welcome and to explain that they are providing a service. This may alleviate test anxieties. They are not the ones being evaluated; the orthography is. They also need reassurance that although they are providing personal information about themselves, their names will not be used and they and their performance results will be kept confidential and unidentifiable in reports.

Several activities should precede the actual test:

*1) A preliminary skill assessment test:* This is necessary to verify that prospective test takers have the requisite skills to participate in the test. Too low a skill level could complicate the analysis and skew test results. For instance, if the test is checking for ease of transfer from the language of education to the local language and requires respondents to match words to pictures, it is important to determine if a prospective participant can interpret pictures and is able to read at the word level. If the test requires reading at the sentence or text level, reading two to three sentences aloud in the language that the test will be conducted in should suffice to determine if the skill level is adequate.  The number of hesitations and errors will serve as indicators. It would be best if the skills assessment does not contain words that are inherently difficult or include the symbolization under investigation. Reading skill is being evaluated at this point, not the orthography. Making audio recordings of the preliminary test is not necessary, but it is necessary to set a skill level standard that determines if a person can participate in the actual test or not.

If a person lacks the requisite skills to participate in the full test, it is important to communicate that they have been helpful. If compensation is indicated for test subjects, those whose performance is too low to be included in the test should not only be thanked, but also compensated for their time.

Participants do not all need to be excellent readers or have the same reading skill level, especially if the *within-subject* research design is used, i.e., one person's reading of one system is compared with the same person's reading of another system. If a group of readers of one writing option is to be compared with readers of another option, participants could potentially be assigned to test groups in a way that assures that groups are fairly equal in skill. Results from the preliminary skills assessment would help with that. Individual reader skills can feed back into the analysis. However, in formal research, controlled assignment of respondents to groups is viewed with suspicion. Only random assignment to groups is acceptable.

*2) Collecting pertinent background information concerning each participant:* This can be done via a questionnaire. If participants are not comfortable filling out forms themselves, an interviewer can do this for them. Frequently such a questionnaire is designed using the language of education since literacy skills in local languages are often not up to par.

3) *Getting evidence of prior informed consent from each participant:* Orthography testing activities constitute "human research." Getting proof of consent is becoming standard practice and often required by academic institutions and funding organizations. Consent to record a participant's voice should be included on forms. (See Appendix D for a sample consent form. Something much simpler may suffice. It might be acceptable to get oral consent from each participant and have one representative sign for all.)

4) *A practice run:* A practice run is considered part of the test, but it is not scored or included in the analysis. It serves as a warm-up exercise and shows participants what to expect, and hopefully, it will relieve any anxiety they might have. If the text comprises different sections, each with different instructions, it will be helpful to include a practice item for each section.

## 5.4    The test

How the actual test is carried out depends, of course, on whether it is a formal or informal test, on the research questions and components of the specific test, and on the skill level of the test takers. Therefore, guidelines presented here may or may not apply in certain situations.

The following test activities are suitable for neo- and semi-literate individuals:

- Matching of words, phrases, or simple sentences to pictures.
- Reading a list of words or individual words on a flashcard series.
- Filling in the missing letter of a word or missing word in a sentence.
- Writing the first sound of a dictated word.
- Reading simple, culture-appropriate sentences.

The description of informal testing done among the Chong [cog] in Thailand (Kosonen 2003) may prove helpful to field workers. Timing lower skilled readers with a stop-watch or recording their readings may not be required. This depends on the research question.

For participants with more advanced literacy skills, the test would usually comprise

- Sentence and text level reading, possibly with alternate orthographies
- Word and sentence level dictation.

Fatigue is likely to affect reading performance on later test components. Researchers need to therefore compensate for this variable by varying the order of texts and the orthographic options these are written in. Although for preference testing and informal initial testing the same text might be presented in two orthographies to the same test subject, this is not acceptable for a formal performance test. Familiarity resulting from the first reading would skew the results for the second reading. Thus when doing tests to check on learnability or efficiency of alternate orthographic options, different texts must be used.

It is important that test designers ascertain in advance that the texts used are natural and fairly equivalent in difficulty, as indicated by sentence length and amount of imbedded discourse. They must not be too difficult; otherwise the exercise is no longer an orthography test but a test of readability and reading skill. Texts should not contain anything outside of cultural norms in terms of content, thus it is best to avoid translated texts. They should not be contrived to have such an abnormal number of occurrences of the feature being tested that they become unnatural or read like a tongue twister.

John Clifton suggests using two texts with readers reading both, but in different orders and in different orthographies to compensate for any differences in text difficulty and fatigue resulting from presentation order (Editor's Postscript to Lauck 1987).

For formal testing, Steve Walter (p.c.) suggests using four texts. Two are to be presented in one orthography (A) and two in the other (B). They can differ in genre, but do not need to. Texts can be

chosen from first and/or third person narratives, fables, or hortative texts, i.e., those trying to affect thought or action, or process texts, i.e., those explaining a procedure, giving step-by-step instructions. *Each* person would read *each* of the four texts (I–IV). Test sets would be prepared in advance so a different order would be presented to the various test subjects, resulting in eight options as presented in Figure 2.

| Test set | 1st text | 2nd text | 3rd text | 4th text |
|---|---|---|---|---|
| 1. | I A | II B | III A | IV B |
| 2. | II B | III A | IV B | I A |
| 3. | III A | IV B | I A | II B |
| 4. | IV B | I A | II B | III A |
| 5. | I B | II A | III B | IV A |
| 6. | II A | III B | IV A | I B |
| 7. | III B | IV A | I B | II A |
| 8. | IV A | I B | II A | III B |

**Figure 2: Variations of text order and orthography options in test sets**

An effort should be made to have approximately the same number of test subjects for each test set. This was done when testing potential orthography options for Belize Kriol [bzj].[8] Decker (2014) describes the test organization and procedure used and, in a table, illustrates how the stories and orthographies were rotated for sixty test subjects. As suggested above, each person read four different stories. However, in the Belize Kriol case, each story was presented in a different orthography, since four options were tested instead of two.

### 5.5 *The test procedure*

Compared to the effort and time that goes into preparation and preliminary activities, the actual test takes relatively little time. It is the advance preparation of staff and tools that helps make this part go smoothly. Guidelines provided here focus on text-level testing since it is the most complex.

As already noted, test subjects who pass the preliminary skills assessment need to be put at ease, receive clear instructions, and have a "warm-up" practice to assure they understand what is expected of them. The following components are part of the real test:

- The test subject reads aloud, one text at a time.
- A voice recording is made of the readings.
- The reading of each text is timed separately.
- The test giver, on a duplicate copy of each text, notes errors and disfluencies such as hesitations, restarts, repetitions, as well as self-corrections of any of these.[9] (Note: This is possible for phrases and short sentences but not usually for texts read by fairly fluent readers. Accuracy would suffer. Recordings will allow this to be done at a later time.)

---

[8] Ethnologue (Lewis, et al., eds., 2014.) reports 70,000 speakers in Belize and a total of 110,000 speakers in the world.

[9] Symbols commonly used by school teachers for lower grade diagnostic reading tests can be used or adapted for orthography testing. (See Appendix E for an explanation of the customary symbols.)

- If a test subject's reading is blocked despite several attempts, the test giver does not intervene, but moves them on to the next text component, if there is one. (Blockages indicate that the person's skill level is inadequate and their results should be excluded from the analysis.)
- After reading the set of texts, the test subject is thanked and asked to do the writing component of the test, if there is one.
- All records pertaining to the test subject need to be stored together, each part marked with the same identification code so nothing is lost. This is important since names can be confusing or similar to one another.

Comprehension questions can optionally be inserted after the reading of each text. This is required if the research question concerns underrepresentation which may result in ambiguity and potential misreadings of homographs. It is, of course, also required if the test consists of silent reading rather than oral reading. Well-crafted questions are essential since they provide data that indicate if underrepresentation is problematic or not. Retelling of a text after oral or silent reading would be another option for testing comprehension.

Test designers might be tempted to deliberately incorporate a number of ambiguous items that are not predictable from context. This especially relates to tone orthographies with partial or no tone notation, which may result in minimal pair lexical or grammatical items which are not easily disambiguated in context. Manipulating texts, and thus the test in this way, artificially increases the functional load of the item being tested and may skew the results. If the texts remain natural, this may be acceptable to a point *if* the investigative question is about determining comprehension of texts with orthographic ambiguity. However, if the investigative question is about something altogether different, such as determining if the orthography facilitates fluent reading in general, or if a certain symbolization works well, there is no need to build in deliberate complexities in texts. A true test of the orthography could otherwise turn into a test for readability of specific artificially contrived texts.

If there are writing components envisioned as part of the test, they can be completed as a group to save time, provided that there is no chance for participants to copy from each other. Activities need to be appropriate for the skill level of test subjects. Those who breeze through the first writing activity can be given additional ones, each increasingly more challenging. It may be wise to set a time limit on completing writing activities.

## 5.6    Number of respondents

There is no hard and fast rule about the size of the sample. However, not having included a larger or a more varied sample in the research is a common critique and regret of researchers. Yet, testing a group of 300 may not reveal anything more than what might have been revealed by testing 70 subjects. The research methodology may affect what would be an appropriate sample size. For instance, for a survey, quite a large number of people can be included without making the process or the analysis of results unmanageable. In contrast, working one-on-one requires limiting the numbers for practical reasons.

Roberts (2008) presents an analysis of sample sizes of ten formal tone orthography testing experiments in West Africa. These ranged from a single subject to sixty participants.  Even a single test subject's reading of texts in a variety of orthographic options can reveal what works and what is problematic symbolization. Roberts does not critique the "within subject" research with a single Kom (bkm) reader (Bernard et al., 1995) but does note that such an approach "is usually reserved for qualitative experiments, particularly, not quantitative ones" (2008: 216). He notes that this experiment was a "starting point," which, indeed, it was, since Bernard's team did a second experiment with the same language group, this time with thirteen Kom subjects (Bernard et al., 2002).

For two of his own experiments, Roberts reports sample sizes of thirty-nine (2010) and fifty-five (Roberts and Walter 2012). He offers the following tips concerning sample size: "Clearly, an initial large

number acts as a kind of insurance against certain unpredictable technical and practical problems, such as missed recordings or unexpected absences. However, high numbers do not necessarily ensure successful results and are not a prerequisite for it" (Roberts 2008:215). Experience shows that it is not just the number of respondents, but the number of conditions that are likely to affect the outcome that determine if results are statistically reliable. Thus, for each variable that is likely to affect outcomes in formal testing, 25-30 test subjects should be included for statistical reliability. (Roberts, p.c.)

An experiment by Bird (1999) illustrates the need to plan for a large enough number. He began with 16 participants, but the number shrank to 11 when a preliminary reading skill assessment indicated that five individuals could identify words but were not able to read texts. Eleven subjects are not deemed an adequate sample size.

Roberts (2008:215) warns that "eliminating subjects can lead to biased results," and, for the other extreme advises, "One should guard against generating excessive amounts of data, because there is a risk of being swamped when it comes to the analysis" because of the amount of data generated. The concern for collecting too much data is mainly an issue for qualitative research which may result in large numbers of pages of data which then need to be examined. Too large a number of respondents and processing findings is not so much an issue in quantitative research since statistics programs can handle large data sets as well as small data sets.

For formal testing, Steve Walter, in training courses, suggests aiming for 60–80 participants. Even if some need to be excluded in the final analysis because of inadequate competence, the chances are good for retaining an adequate number for research findings to be proven statistically significant. A group of 70, even if divided up for inter-group comparison, would leave the control group and experimental group sizeable enough to not draw criticism or risk loss of validity.

## 5.7    *Analysis of the data and reporting*

Findings will need to be examined thoroughly and organized in a variety of ways to reveal significant factors and correlations. The data, in various organized formats, should indicate how the various orthography options fared in acceptance and performance. If there is a measureable difference between options or test groups, one needs to find out if the difference is significant or might have happened by chance. A Chi-square test or a t-test can be used to determine this if only two test groups are involved. For more complex tests different kinds of statistical tests are required.

Spreadsheets and calculators have shortcut functions to provide the necessary figures to determine statistical significance. But determining what kind of test is appropriate for which situation takes specialized knowledge. It is not recommended that the person planning and implementing orthography testing also take charge of the analysis of the data unless they were specifically trained in and have experience in research design and statistical analysis. It is best to call on a statistician for help not only with the analysis, but also the interpretation of the findings and reporting.

Analysis can reveal if differences in findings correlate with specific independent variables. To discover such correlations, results need to be examined as a whole and also in sub-groupings based on the various background factors. This could indicate differences in the data based on gender, age, level of education, dialect, additional languages spoken or written, amount of time spent outside of the language area, etc.

Analyzing quantitative test results will be more straightforward than analyzing qualitative findings since the results are already in a format that allows mathematical calculations. Qualitative data needs to first be coded quantitatively before statistical analysis can be done.

Displaying data in ways that facilitate the analytical process will be helpful to the researcher and likely lead to appropriate conclusions. It is important to not lose sight of what the original research questions were and always relate findings back to these.

Readers of reports will appreciate the inclusion of helpful visuals since this will eliminate wading through excessive prose. Print journals, on the other hand, to economize on space, may not welcome a large number of figures and tables.

## 6. Characteristics of formal testing

As already noted, what makes the formal research situation different from an informal one is not necessarily the investigative questions, the amount of rigor put into the testing, or the testing activities. Certain types of activities can serve both types of testing. There are, however, some differentiating characteristics.[10]

### 6.1 Required formalism

Scientific research needs to be objective and adhere to certain design regulations. One of the unique features of formal testing is the specification of two logical statements expressed as two different hypotheses: *a null hypotheses* and an *alternative hypothesis*. This is part of the scientific method. The logical statements need to adhere to very specific rules:

- The null hypothesis always contains an equals sign. (Options are $=, \leq$, or $\geq$, which stand for "equal to," "less than or equal to," and "greater than or equal to.")
- The alternative hypothesis is the logical converse of the null hypothesis. It never includes a statement containing an equal sign if expressed in logical shorthand. (Only $>, <$, or $\neq$ may be used, which stand for "greater than," "less than," or "not equal to.")

The second part of the formalism is a *claim*. The claim indicates which of the two logical statements or hypotheses the researcher believes to be true. The claim may be associated with either the null or the alternative hypothesis, depending on what the researcher's position or belief happens to be. However, in formal research, it is **always** the null hypothesis whose truth value is tested by the statistical analysis that follows. If the findings do not support the null hypothesis, it is rejected and one concludes, by logical implication, that the data then provide support for the alternative hypothesis. If the researcher's claim is associated with the null hypothesis, and the data fail to support the null hypothesis, the claim will be rejected as well. If the claim is associated with the alternative hypothesis, a rejection of the null hypothesis is interpreted as support for the alternative hypothesis.

The following examples illustrate investigative questions with potential associated paired alternative and null hypotheses and discuss what might influence which one the researcher might hold to:

1. *Investigative question:* "Will symbolizing /ŋ/ as <ng> while symbolizing nasalized vowels as vowels followed by <n> interfere with reading fluency due to ambiguity as to whether <n> stands for nasalization of the previous vowel or the first element of the <ng> digraph?"
   a) *Alternative hypothesis:* Reading fluency performance will be inferior for texts which symbolize /ŋ/ as <ng>, as opposed to <n͡g>.
   b) *Null hypothesis:* Reading fluency will be the same or better for texts in which /ŋ/ is written as <ng> compared to those in which /ŋ/ is symbolized by <n͡g>.[11]
   *Claim:* If resultant ambiguity in texts is expected to be high, researchers would most likely associate the claim with the alternative hypothesis. If not, their claim would likely be associated with the null hypothesis, stating that the <ng> symbolization would not be problematic and that there would be no significant difference in performance.

2. *Investigative question:* "Will marking lexical tone in addition to grammatical tone improve reading fluency and reading comprehension in language XYZ?"

---

[10] I am grateful to Dr. Steve Walter for his assistance with section 6 of this paper.

[11] Framing the null hypothesis in this way means that <n͡g> is the only orthographic alternative being tested and compared with <ng>. A three-way comparison could also be made, but would require a more complex test design.

a) *Alternative hypothesis:* Readers will demonstrate superior reading fluency and comprehension with XYZ texts which mark lexical tone in addition to grammatical tone as compared to performance on texts which only have grammatical tone marked.
b) *Null hypothesis:* Reading fluency and comprehension will be the same or inferior for XYZ texts which mark lexical tone in addition to grammatical tone compared to texts which mark lexical tone only.
*Claim:* If ambiguity in texts is expected to be high due to high functional load of lexical tone, researchers would most likely associate the claim with the alternative hypothesis. If not, their claim would likely be associated with the null hypothesis, stating that noting lexical tone will not result in additional benefit for reading fluency or comprehension. (Note that the research question does not specify how or how much lexical tone would be marked.)

In both examples, the researcher can associate the claim (i.e., "stated belief") with either hypothesis. Whether the resultant data supports the researcher's claim depends on the data itself and which hypothesis the claim is associated with.

## 6.2    The analysis

Statistical rigor is a must in formal analysis. It bears repeating that researchers not trained in statistical methods would do well to ask for help from someone who is. Researchers need to resist the temptation to exclude puzzling findings from the report. These may provide insight at a later date.

Besides analyzing the data in terms of the main research question, a researcher may also wish to discover if correlations exist between the dependent variables ( i.e., the items one is testing and measuring, such as fluency, number of errors, comprehension) and the independent variables (the known factors which can be controlled, such as age, gender, educational level, etc. of the respondents). For instance, do higher numbers of reading errors correlate with certain dialects? Do the symbolization preferences correlate with test participants' religious affiliation, their age, or educational level? Specialized analytical tests can pinpoint such correlations based on information provided on background questionnaires.

Decker reports on an interesting correlation that came to light during a formal orthography performance test for Belize Kriol. Analysis revealed a correlation between subjects' performance toward the end of the reading test and how they had answered the question, "What do you think of the idea of reading and writing Kriol?" Decker found that individuals either improved in skill, i.e., read later stories faster, or, else slowed down and made more mistakes: "If they felt positive towards the idea of literacy in Kriol, they tended to be the ones who had improved. If they felt negative or ambivalent towards reading and writing in Kriol, they tended to tire of the exercise" (2014:161).

## 6.3    Tests for statistical significance

Tests do not always yield data according to expectations even with tight controls in place and few variables. For instance, a coin flip test does not usually turn out exactly 50% heads and 50% tails. It is necessary to determine if deviations from expected results might have happened by chance, or if the observed difference seems to be statistically significant. Chi-square tests and t-tests are useful tools to verify statistical significance.

A Chi-square test is particularly helpful in analysis of attitudinal research. It is usually applied when the data relates to a *counting* indicator. For example, to determine if a language group prefers one symbol choice over another, the Chi-square test is used to ascertain if the difference in preference within the respondent sample is significant or might have happened by chance. This is the simplest type of Chi-square test, often referred to as the *goodness-of-fit* test. If the observed difference proves to be statistically significant, the null hypothesis must be rejected. The *goodness-of-fit* test applies when there is a focus on one variable. For example, "Did all the teachers tend to answer a certain question the same way?"

A second type of Chi-square test is known as a *test of independence* of two populations. It answers the question whether or not a subgroup of the sample population has the same results as another subgroup, or if there is a significant difference in their responses? For example, did women and men tend to respond similarly to specific questions? A *test of independence* needs to be used when comparing data involving two variables or subgroups.

One major factor that plays into the calculations and statistical significance is the sample size. The researcher needs to be aware of conditions that may require an increase in sample size. For instance, on the surface, a survey for symbolization preference and one checking on script preference seem to provide a similar type of data. However, identity plays a more significant role in script choice. Since religious affiliation and/or living in proximity to a neighboring country in the case of a cross-border language situation are factors that may greatly impact results, researchers need to assure adequate background variety in the sample in addition to aiming for an adequate sample size.

A t-test is quite different from a Chi-square test. It is usually applied to test for statistical significance when the data represents a *measure* of some kind. For example, the number of reading errors made in texts written in different orthographic traditions, or how long people took to read various texts, would provide measures that would lend themselves to verification via a t-test.

The scope of this paper does not permit expansion on this vast topic.

## 7. Conclusion

Orthography testing may seem like an extraneous and extravagant undertaking to those who consider orthography development to be a fairly straightforward process. But experience indicates that "[i]t is not enough to propose an alphabet based on linguistic analysis and socio-cultural factors, it must be field-tested to discover how adequate it is, how easy it is to read and to write and what kinds of problems may emerge" (Robinson & Gadelii 2003:40). So, in the context of sociolinguistic realities, linguistic complexities, and limited opportunities for the use of certain languages in education, it is wise to plan for testing a newly developed or revised orthography. This will facilitate the language's development and, hopefully, promote its use and survival in situations where speakers have several language choices for communication and learning. An efficient orthography, if embraced by the speech community, will encourage language use in the written domain and may open up additional possibilities for its use in formal and non-formal education.

Orthography testing will not look the same for all languages since different investigative questions need answering. Some investigative questions will relate to preferences in symbolization and acceptability of the writing system as a whole. Others will relate to the comparative efficiency of specific symbolization options or user-friendliness of the system. Informal testing will be adequate in most cases; in others, formal testing is the better choice. Good planning, well-designed tools and careful analysis should characterize all testing. Researchers are urged to take the time to report on their research and findings so that, in future, others involved in language development activities may also benefit.

This article is my attempt to provide some practical, accessible guidelines for orthography testing since, to date, there has been a void in this domain. There is no "one-size-fits-all" plan. But some practices are better than others, and I have attempted to highlight these. Yes, doing at least some orthography testing will surely be better than doing none. However, it's not about meeting expectations or minimal requirements, but about making every effort to ensure that end-users will be well-served.

## APPENDIX A: The Background Questionnaire

This is a basic background questionnaire that would supply helpful information for orthography test result analysis and determining correlations between findings and independent variables.

---

**Personal Information:**

**Name:** _____          **Male:**   ☐

**Place of birth:** _____ # of years _____          **Female:**  ☐

**Place of residence:** _____ # of years ____          **Age:**  ___

**What languages do you speak?** _____

**What languages can you read?** _____

**What languages can you write?** _____

**Language(s) spoken at home by mother:** _____

**Language(s) spoken at home by father:**  _____

**Level of education completed:**
☐  non-formal literacy classes
☐  1-3 yrs. of primary
☐  all of primary
☐  some of secondary
☐  all of secondary
☐  university
degree:  yes ☐      no ☐

**Profession:**  _____          **Religious affiliation:** _____

**How often do you read per week?** _____  **What?** _____

**In which language(s)** _____

**Individual Code assigned:** _____

---

**Note:** The education section could be simplified to **"What is the last year of schooling you completed?"** This basic template should be supplemented with context-appropriate questions. For instance,

- If the language of instruction is not the same in all schools, it would be appropriate to ask, "In what language did you first learn to read?" and "What language(s) did your first teacher use in class?"
- If places of worship vary in language use or in orthographies employed in their literature, questions revealing such influencing factors should be included.
- Basic questions that reveal language attitudes that might affect orthography test results are always appropriate, such as, "Which language do you speak to your children most of the time?" or "Do you think the W_____ language should be used to teach reading and writing in school?"

## APPENDIX B: Inviting Feedback on a Tentative Orthography

The following are some suggested approaches for inviting feedback on a tentative orthography. Which type of response mechanisms are appropriate will depend on the degree of isolation or dispersion of the language community and the communication technologies available to them.

### *Situation 1: Radio or television broadcast in the local language, with announcement*

Thank you for listening to _____ (title). We hope you have enjoyed this program in the W_____ language. It is based on the story _____ (title), written by _____ (author). It is available in print from various market vendors and in shops for only _____ (price).

The W_____ language has been written in a variety of ways these past ____ years. How to best write it is still under discussion. We would like your feedback on how this drama/story looks to you in print. For instance, what do you think about the current suggested way of writing W_____ ? What do you like about it? Is it easy or hard for you to read? What would you like to see change?

Please call or text _____ (phone #) to pass on your comments to the W_____ Language Committee, Or, write the Committee, care of this station. Or: talk to any W_____ literacy teacher in the region. We and they will make sure the Committee receives your comments.

### *Situation 2: Front matter in a publication*

Thank you for your interest in literature in the W_____ language. We hope you will enjoy this book/will find this resource helpful.

The W_____ language has been written in a variety of ways these past ____ years. How to best write it is still under discussion. Here are the symbols with illustrative words and brief sentences that show the current suggested way of writing W_____ .

*[grapheme list with a key word and sentence for each here]*

We value your opinion about this way of writing the W_____ language. We would like to know what you like about it. Tell us if you think it easy or hard for you to read, and if there is anything you would like to see changed.
1. Please come visit us at the W_____ Language Committee office at _____ (location) or
2. write us a note with your reaction, or
3. call or text us at _____ (phone number), or
4. talk to any W_____ literacy teacher in the region.

### *Situation 3: Insert in publications*

Thank you for your interest in literature in the W_____ language. After you have looked through this book/used this resource, please fill in this survey questionnaire.

We value your opinion about this way of writing the W_____ language. How to best write it is still under discussion. We would like to know what you like about it and whether it easy or hard for you to read, and if there is anything you would like to see changed.

Please fill out this form and come visit us at the W_____ Language Committee office at _____ (location) or give the form to any W_____ literacy teacher in the region.
All information and your feedback will be kept confidential. Thank you.

*[Orthography survey questions would follow. These would be specific to the complexities of the language in question and the investigative questions that need to be answered.]*

## APPENDIX C: Informal Preference Testing

Preference testing can be done at the word, phrase, sentence or text level. Varying the order is important. Response forms should be accompanied by respondent's background information so correlations between answers and independent factors can be discovered.

### 1. Word level testing instructions:
Please listen to each word I will read to you. Look at the options of how that word could be written. Choose the option you like best and underline it.

|    | A | B | C | D |
|----|------|------|------|------|
| 1. | fait | fite | fiht | fyt |
| 2. | tipe | typ | taip | tihp |
| 3. | rihd | raid | ryd | ride |
| 4. | ys | ihs | ise | ais |
| 5. | drie | drai | drih | dry |
| | etc. | | | |

### 2. Sentence level testing instructions:
Please listen to the sentences I will read to you. Look at the different ways those sentences are written. Choose the option you like best and underline it.

| 1. | The sun was schining so he went fisching. |
|----|------|
| | The sun was ʃining so he went fiʃing. |
| | The sun was šining, so he went fišing. |
| | The sun was s̱ining, so he went fis̱ing. |
| 2. | He could not buy the ʃovel because he didn't have enough caʃ. |
| | He could not buy the s̱ovel because he didn't have enough cas̱. |
| | He could not buy the schovel because he didn't have enough casch. |
| | He could not buy the šovel because he didn't have enough caš. |
| 3. | I hope the s̱eep will not eat the leaves off the bus̱. |
| | I hope the šeep will not eat the leaves off the buš. |
| | I hope the ʃeep will not eat the leaves off the buʃ. |
| | I hope the scheep will not eat the leaves off the busch. |
| | etc. |

### 3 a) Text level testing: Script preference testing
(Presentation order to be reversed for half of the respondents.)

> …Tongaso lakue na lakui, lo yeke gue na ya ti da ti awali, amama kondo ni, oko na oko ti baa ala. …
> Gi na lege ti nduru bungbingo tere ti ala so, Baambi awara sioni kobela so. …
> *(The complete story should appear. For illustrative purposes these two sentences suffice.)*

> Тонгасо лакуе на лакуи, ло ыеке гуе на ыа ти да ти авали, амама кондо ни, око на око ти баа ала. …
> Ги на леге ти ндуру бунгбинго тере ти ала со, Баамби авара сиони кобела со. …[12]

> *Suggested instructions and questions:*
> Please look at these different ways of writing the same story.
>
> 1. Which way of writing this story do you like better?
> 2. Can you read both versions of the story? Please read.
>    Which way of writing seems easier to read?
> 3. If the _____ language were to be taught in schools and text books are developed, which writing system do you think would be a better choice? Why?
>    *Optional perception questions:*
> a) Do you think other people would like one system better than the other?
> b) Who might like the first way of writing better? Why?
> c) Who might like the second way of writing better? Why?

### 3 b) Text level testing: Determining individuals' preferences between two symbolization options
(Testing two labialization options; presentation order to be reversed for half of the respondents.)

> …Tongaso lakue na lakui, lo yeke gue na ya ti da ti awali, amama kondo ni, oko na oko ti baa ala. …
> Lo nyon yoro gba si Baambi, pendere koli kondo so, akui. Fade e toto kua ti Baambi so tongana nye? Sioni pupu asi na ya ti kodoro ni. Kua ahondoni. …
> *(The complete story should appear. For illustrative purposes these sentences suffice.)*

> … Tongaso lakwe na lakwi, lo yeke gwe na ya ti da ti awali, amama kondo ni, oko na oko ti baa ala. …
> Lo nyon yoro gba si Baambi, pendere koli kondo so, akwi. Fade e toto kwa ti Baambi so tongana nye? Sioni pupu asi na ya ti kodoro ni. Kwa ahondoni. …

> *Suggested instructions and questions:*
> Please look at these different ways of writing the same story. Please read both through out loud and tell me what you think. When you are finished reading I will ask some questions:
>
> 1. Which version of the story was easier to read? Why do you think that is?
> 2. Which words seemed difficult?
> 3. Was there anything that seemed difficult in both stories?

---

[12] Credit for this Latin to Cyrillic transcription goes to *Lexilogos* multilingual keyboard.

*3 c) Text level testing: Determining individuals' preferences between writing systems*
(Comparison of two or more systems of writing involving multiple symbolization differences. Presentation order is to be varied so each variation of the text is presented in 1st, 2nd or 3rd position approximately the same number of times. For formal testing different stories would be used.)

> … Tongaso lakue na lakui, lo yeke gue na ya ti da ti awali, amama kondo ni, oko na oko ti baa ala. …
> Lo nyon yoro gba si Baambi, pendere koli kondo so, akui. Fade e toto kua ti Baambi so tongana nye? Sioni pupu asi na ya ti kodoro ni. Kua ahondoni.  …
> *(The complete story should appear. For illustrative purposes these sentences suffice.)*

> … Tongaso lakoue na lakoui, lo yeke goue na ya ti da ti aouali, amama kondo ni, oko na oko ti ba ala. …
> Lo gno yoro gba si Bambi, pendere koli kondo so, akoui. Fade e toto koua ti Bambi so tongana gne? Sioni poupou assi na ya ti kodro ni. Koua ahondoni.  …

> … Töngasô lâkûê na lâkûi, lo yeke gue na yâ tî da tî âwalï, âmamâ kôndo nî, ôko na ôko tî bâa âla. …
> Lo nyön yorö gbä sï Bâambï, pendere kôlï kôndo sô, akûi. Fadë ë toto kûâ ti Bâambï sô töngana nye? Sïönî pupu asï na yâ tî ködörö nî. Kûâ ahöndönî. … [13]

> *Suggested instructions and questions:*
> Please look at these different ways of writing the same story. Please read them through out loud. I'd like to know what you think about each.
> When you are finished reading I will ask you some questions:
> 1. Which way of writing the story did you like best?
> 2. Which way of writing seemed easiest to read?
> 3. Which way of writing seemed hardest to read?
> 4. Which words seemed the most difficult?
> 5. Was there anything that seemed difficult or easy in all versions of the story?
> 6. Do you have any other comments about any of the spellings?

---

[13] This excerpt represents the Sango orthography which this story, *Pakara Bâambï* (*Mr. Look-At-Me*, by Karan & Moehama 1993), was actually published in. Diacritics represent tone. The various other excerpts presented in Appendix C were modified for illustrative purposes. The Latin script spellings which appear are representative of various orthographic variations in use for Sango in the 1990s.

The conversion into Cyrillic script encountered a glitch since Sango consonants <w, y> have no Cyrillic equivalents.

## APPENDIX D: Sample Consent Form

Certain human subjects research may be eligible for "exempt" status. Appropriate paper work needs to be submitted to an Institutional Review Board office to assure compliance and integrity in research.

The most common requirement for human subject research is the gathering of prior informed consent. There is no single format that must be followed for a consent form, but certain elements must be present. What is important is that test subjects understand their rights, the procedure and purpose of the test, and that they are reassured that they will not be identified by name in reports. The form below is not to be used as a template. It simply illustrates common components of a consent form for human subject testing. An abbreviated form may be acceptable. If the researcher is affiliated with a specific university, and hopes to publish with them, it is best to access their specific documents, guidelines and templates.[14]

---

**Consent to Participate in Research**

*W_____ Orthography Testing*

*Project Director: _____    Organization: _____    Ministry of _____    Permit #:____*

**Statement of Research:**

A person who is to participate in research must agree to do so of their own free will. Consent must be based on understanding the research. You may choose to participate or not to participate. Please take your time in deciding. If you have any questions, please ask.

**What is the purpose of this research?**

You are invited to help researchers discover the best way to write the W____ language.

**How many people will participate?**

Approximately 12 people will participate here in this location. There will be a total of 5 testing sites, so approximately 60 people in total will participate.

**How long will this take?**

Your participation in the actual activities will take about one hour. But waiting for your turn might take an additional one or two hours.

**What will happen during the research?**

Researchers will ask questions that will help them discover what symbols people prefer. You will be asked to do some reading and writing exercises which will help researchers find out which way of writing W____ might work better. Every task will be explained to you.

**What are the risks of me participating in this?**

There are no foreseeable risks in participating. However, some questions might make you uncomfortable. If that's the case you can choose not to answer. Being asked to read and write texts in W_____ might make you nervous or frustrated. Remember that this is not a test of your abilities. What is being tested and compared is different ways of writing W___ , and you are being helpful. You can change your mind and stop participating at any point.

---

[14] The sample presented here is based on elements included in the sample consent form provided online by the Institutional Review Board of the University of North Dakota.

**What are the benefits of this research?**

Based on what researchers discover, W____ will be written in a way that will make reading and writing W_____ as easy as possible. This will motivate reading and writing in W_____.

**Will it cost me anything to participate?**

You will not have any expenses for participating, but will sacrifice some of your time.

**Will I be paid for participating?**

You will not be paid for being in this research study. However, _____ will be given you to cover the cost of a drink and snack.

**Who is funding this research?**

No funding has been received for this specific research. It is an integral part of the W_____ language development project supervised by _____ and the W _____ language committee.

**Confidentiality**

All records of this study will be kept confidential. In reports or publications you will not be identified. Instead, a summary of the findings will be provided. When referring to individuals, codes, not names will be used. Please initial the following, if you agree:

**I give consent to be audiotaped during this session.** _____Yes _____ No

**I give consent to be photographed.** _____Yes _____ No

**I give consent for things I say to be quoted, without identifying me. _____Yes _____ No**


Your signature below indicates that the research you are invited to participate in has been explained to you, that you had the opportunity to ask questions and that your questions have been answered.

Your signature indicates that you agree to participate in this research.

Subject's Name: _____

Signature of Subject: _____

Location: _____  Date: _____

Parent/Guardian/School official signature if subject is a minor: _____


I have discussed the above points with the subject, or where appropriate, with the subject's legally authorized representative.

_____          _____

   Signature of Person Who Obtained Consent                    date

**APPENDIX E: Oral Reading Error Analysis**

Reading error analysis is commonly used to identify which reading strategies a person is using or not using. The analysis seeks to discover whether or not the reader is strong or weak in making use of the grapho/phonic, semantic, and syntactic systems and serves as the basis for intervention if weaknesses are identified. Current practices are rooted in the work on *miscue analysis* by Kenneth Goodman (1969).

In orthography testing the reader is not in focus. But the expectation is that reader errors will help identify weaknesses in the orthographic code. The following notation system is commonly used to note errors: (Learning and Skills Improvement Service 2014.)

1. **Substitution:** the word substituted is written above the word that should have been read.

   *pills*
   ex.    She was upset because she couldn't find her bills.

2. **Self-correction:**

   *pills©*
   ex. She was upset because she couldn't find her bills.

3. **Mis-correction**: (first read wrong as *pills*, then *balls*)

   *pills mc*
   ex. She was upset because she couldn't find her bills.

4. **Repetition:** the word or words repeated are underlined.

   ex. She was upset <u>because she couldn't</u> find her bills.

5. **Omission:** word(s) left out are circled.

   ex. She was upset (because) she couldn't find her bills.

6. **Insertion:** word(s) inserted are written above with caret indicating the place.

   *very*
   ex. She was ^upset because she couldn't find her bills.

7. **Reversal:** continuous line for what was reversed.

   ex. She got better quickly after resting for a while.

8. **Hesitation or long pause:**

   ex. She got better/quickly after resting//for a while.

9. **Long pause, requiring help:** (^T notes teacher intervention)[15]

   ex. She got better//^T quickly after resting for a while.

10. **Refusal, non-response:** dotted line indicates section not read.

    ex. She got better quickly <u>after resting for a while</u>.

---

[15] Intervention is not recommended in orthography testing or for standardized reading tests.

**References:**

Berkowitz, Susan. 1997. Analyzing Qualitative Data. Chapter 4 in Joy Frechtling and Laure Sharp (eds.), *User-Friendly Handbook for Mixed Method Evaluations*. National Science Foundation, Division of Research, Evaluation and Communication, Directorate for Education and Human Resources. http://www.nsf.gov/pubs/1997/nsf97153/chap_4.htm  Accessed November 21, 2014.

Bernard, H. Russell, George Ngong Mbeh and W. Penn Handwerker. 1995. The tone problem. In: A. Traill, R. Vossen, and M. Biesele (eds.), *The Complete Linguist* — Papers in Memory of Patrick J. Dickens. Cologne: Rüdiger Köppe, 27–44, cited in Bird 1999.

Bernard, H. Russell, George Ngong Mbeh and W. Penn Handwerker. 2002. Does marking tone make tone languages easier to read? *Human Organization* 61(4), 339–349. http://nersp.osg.ufl.edu/~ufruss/publications.htm  Accessed November 21, 2014.

Berry, Jack. 1977. 'The Making of Alphabets' revisited. In: Joshua. A. Fishman (ed.), *Advances in the Creation and Revision of Writing Systems*. The Hague: Mouton, 3–16.

Bird, Steven. 1999. When marking tone reduces fluency: An orthography experiment in Cameroon. *Language and Speech* 42(1), 83–115.

Cahill, Michael, and Elke Karan. 2008. Factors in Designing Effective Orthographies for Unwritten Languages. SIL Electronic Working Papers. 2008-001. SIL International. http://www-01.sil.org/silewp/2008/silewp2008-001.pdf  Accessed Sept. 12, 2013.

David and Lucile Packard Foundation. 2014. Evaluation Philosophy. http://www.packard.org/about-the-foundation/how-we-operate-2/evaluation/evaluation-philosophy/ Accessed March 19, 2014.

Decker, Kendall Don. 2014. *Orthography Development for Creole Languages*. PhD Thesis. University of Groningen. http://dissertations.ub.rug.nl/faculties/arts/2014/k.d.decker/ ?pLanguage=en&pFullItemRecord=ON Accessed December 23, 2014.

Goodman, Kenneth S., 1969. Analysis of oral reading miscues: Applied psycholinguistics. *Reading Reasearch Quarterly* 5, 9–30.

Hinton, Leanne. 2014. Orthography Wars. In Michael Cahill and Keren Rice (eds.), *Developing Orthographies for Unwritten Languages,* 139–168.

Karan, Elke. 2006. Writing system development and reform: A process. M.A. thesis. University of North Dakota.  http://arts-sciences.und.edu/summer-institute-of-linguistics/theses/_files/docs/2006-karan-elke.pdf Accessed November 21, 2014.

Karan, Elke. 2013. The ABD of orthography testing: Integrating formal or informal testing into the orthography development process. Work Papers of the Summer Institute of Linguistics, University of North Dakota Session, vol. 53. http://arts-sciences.und.edu/summer-institute-of-linguistics/work-papers/_files/docs/2013-karan.pdf  Accessed November 21, 2014.

Karan, Elke. 2014. Standardization: What's the Hurry? In Michael Cahill and Keren Rice (eds.), *Developing Orthographies for Unwritten Languages,* 107–138.

Karan, Elke and Elisée Moehama. 1993. Pakara Bâambï (Short Story: Mr. Look-at-Me). In Gä terê ôko. Bangui, Central African Republic: ACATBA.

Karan, Mark. 2009. A Simple Introduction to Qualitative Data and its Qualitative or Quantitative Analysis. Unpublished Manuscript.

Kosonen, Kimmo. 2003. Testing Chong orthography in Thailand. Paper presented at the 1st Int'l Conference on Language Development, Revitalization and Multilingual Education in Minority Communities in Asia. 6–8 November 2003.  Bankok: Institute of Language and Culture for Rural Development, Mahidol University. http://www.sil.org/asia/ldc/parallel_presentations.html Accessed November 21, 2014.

Lauck, Linda M. 1987. Orthography testing in Patep. In: John M. Clifton (ed.), *Studies in Melanesian orthographies*. Data Papers on Papua New Guinea Languages 33. Ukarumpa: SIL, 151–159. http://www-01.sil.org/pacific/png/pubs/928474523780/Orthography_Testing_Patep.pdf Accessed October 25, 2013.

Learning and Skills Improvement Service. 2014. Miscue Analysis. Diagnostic assessment: literacy: materials for assessing reading using miscue analysis. Read Write Plus. Excellence Gateway. http://rwp.excellencegateway.org.uk/resource/Diagnostic+assessment%3A+literacy%3A+materials+for+assessing+reading+using+miscue+analysis/pdf/Accessed February 24, 2015.

Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig. 2014 (eds.). Ethnologue: Languages of the world, Seventeenth edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com  Accessed November 21, 2014.

Lexilogos. 2002–2014.  *Lexilogos: Words and Wonders of the World.* Multilingual Keyboard: Cyrillic-Latin conversion.  http://www.lexilogos.com/keyboard/russian_conversion.htm . Accessed November 21, 2014.

Markowski, Linda. 2009. So Orthography Committee and Revision Process. Working Paper. Chiang Mai: Payap University. http://ic.payap.ac.th/graduate/linguistics/papers/ So_Orthography_Development.pdf Accessed November 21, 2014.

Moe, Ronald. 2005. Developing an orthography for a Bantu Language. Unpublished manuscript.

Pell Institute and Pathways to College Network. 2014. Evaluation Toolkit: Evaluation Myths and Benefits. The Pell Institute for the Study of Opportunity in Higher Education, the Institute for Higher Education Policy, and Pathways to College.  http://toolkit.pellinstitute.org/evaluation-101/evaluation-myths-benefits/ Accessed March 19, 2014.

Powlison, Paul S. 1968. Bases for formulating an efficient orthography. *The Bible Translator* 19 (2), 74–91.

Roberts, David. 2008. Thirty years of tone orthography testing in West African languages (1977–2007). *Journal of West African Languages* 35 (1-2), 199–242.

Roberts, David. 2010. Hidden morpheme boundaries in Kabiye: A source of miscues in a toneless orthography. *Writing Systems Research* 2:2, 139–153.

Roberts, David. 2011. Autosegmental and pedagogical considerations in preparation for a tone experiment. *Journal of West African Languages* 38 (2), 87–106.

Roberts, David, and Stephen L. Walter. 2012. Writing grammar rather than tone — an orthography experiment in Togo. *Written Language and Literacy* 15(2), 226–253.

Roberts, David, Keith Snider and Steven L. Walter. (forthcoming). Neither deep nor shallow: testing the optimal orthographic depth for the representation of tone in Kabiye (Togo).

Robinson, Clinton and Karl Gadelii. 2003. Writing Unwritten Languages: A Guide to the Process. Working paper. UNESCO. http://unesdoc.unesco.org/images/0022/002264/226475E.pdf Accessed November 21, 2014.

Smalley, William A. 1959. How shall I write this language? *The Bible Translator* 10(2), 49–69. Reprinted in Smalley et al., 31–52.

Smalley, William A. 1964. Writing Systems and Their Characteristics. In William A. Smalley (ed.), *Orthography Studies: Articles on New Writing Systems.* Helps for Translators 6. London: United Bible Societies, 1–17.

Smalley, William A. (ed.). 1964. *Orthography studies: Articles on new writing systems.* Helps for Translators 6. London: United Bible Societies.

Snider, Keith L. 2001. Linguistic Factors in Orthography Design. In: Ngessimo M. Mutaka and Sammy B. Chumbow (eds.), *Research Mate in African Linguistics: Focus on Cameroon*. Köln: Rüdiger Köppe Verlag, 323–332.

Snider, Keith L. 2014a. Orthography and Phonological Depth. In Michael Cahill and Keren Rice (eds.), *Developing Orthographies for Unwritten Languages*, 27–48.

Snider, Keith L. 2014b. *Tone Analysis for Field Linguists.* Ms., SIL International.

Trochim, William M.K. 2006. The Qualitative – Quantitative Debate. Social Research Methods. http://www.socialresearchmethods.net/kb/qualdeb.php  Accessed November 21, 2014.

University of North Dakota. 2014. Informed Consent Document Template: Non-Medical Projects. Institutional Review Board. UND. http://und.edu/research/resources/human-subjects/forms.cfm Accessed December 1, 2014.

Venezky, Richard L. 1970. Principles for the Design of Practical Writing Systems. *Anthropological Linguistics* 12(7), 256–270. Reprinted in 1977: Joshua. A. Fishman (ed.), *Advances in the Creation and Revision of Writing Systems*. The Hague: Mouton, 37–54.

Walker, Willard. 1969. Notes on Native Writing Systems and the design of native literacy programs. *Anthropological Linguistics*. 11(5), 148–166.