

Pilot Certification Testing of Health Care Interpreters: Lessons Learned and Recommendations

Elizabeth Nguyen, California Healthcare Interpreters Association (CHIA)

I am here to represent CHIA as a board member and former co-chair of the CHIA Standards and Certification committee.

The California Healthcare Interpreters Association is a 501-C3 public charity dedicated to serving the public good. Our mission is health care interpreters and providers working together to overcome linguistic and cultural barriers to high quality care.

I was asked to share the findings and lessons learned from the joint pilot held in 2002-2003. This was a collaborative project between Massachusetts Medical Interpreters Association (MMIA) and California Health Care Interpreting Association (CHIA) to test the prototype of a certification exam created by the MMIA (now IMIA). The pilot was made possible through funding provided by the Office of Minority Health to the National Council on Interpreting in Health Care (NCIHC).

The purpose of the joint project was for NCIHC to support regional efforts, for the MMIA to obtain a larger sample size for its testing prototype, and for CHIA to learn lessons that will guide the recommendations for future certification.

CHIA's participation in the joint pilot was to assist the MMIA in the selection and training of the test administrators, proctors, and raters; to recruit and identify the test candidates; to support the coordination and preparation process; to assist in the implementation; and at the end of the testing, to discuss and provide some of the feedback. I would like to clarify the role of the CHIA Standards and Certification Committee (S&CC) during that time, and in relation to the pilot testing. At the time of the joint project, we had just developed and published the CHIA *California Standards of Practice for Health Care Interpreters* in September 2002, with funding from The California Endowment. While we were interested in gaining experience with the pilot testing, the S&C committee was not at all involved in the design of the testing, nor in determining the validity or reliability of that testing tool.

About the testing tool: the IMIA medical interpreter assessment certification tool has four modules. In module number one, there is a written test regarding medical terminology. In module two, there is another written test that has to do with standards related to ethical and cultural issues. Module three covers language conversion from Spanish to English and English to Spanish. And, of course, there is a role play that would test the testers on how they would perform in a certain situation.

We proceeded with selecting the administrators, and these are some of the basic characteristics identified in choosing them. They needed to have three or more years of experience as paid medical interpreters. They needed to be respected by others in their field. So we asked for references. They had to have experience training or supervising interpreters. And of course, they had to have a commitment and a passion for the field and for the development of the profession. And we asked that they also engage in ongoing professional development.

Some specifics about the administrators in California: we had thirteen people. And the countries that they represented ranged from Argentina all the way to Spain. They are highly experienced interpreters. They have worked in their field at least ten years. They have diverse educational experience and background. They are aware of Spanish regional differences and they are quite sensitive to language issues among heritage speakers. We are in California where there are about 334 languages represented. And within the Spanish language alone, there are many regionalisms, so identifying the characteristics of these administrators was very important.

How did we recruit the participants? Our goal was to identify participants who represent three levels of proficiency, from “beginner” all the way to “advanced”. They had to fill out a questionnaire where we asked them about their employment history as medical interpreters. At that time, we also included people who just said, “Well I just volunteer as an interpreter.” We also asked about the type and the amount of specific training or education they had received. This is self-declared. Also, we asked them about the estimated number of interpreter encounters or the numbers of hours they had worked as interpreters. So it was a pretty basic questionnaire.

We conducted this pilot in partnership with Healthy House. Healthy House is a community-based organization, based in Merced, California, in the Central Valley. We had forty-six participants representing California. This is how we prepared the candidates for the test. We told them the essential readings that they had to have, such as the *California Standards for Health Care Interpreters*. Bear in mind that, at the time the pilot testing started, we had just published the Standards book. It was just September 2002 when we published it, and so we didn’t have the time to go out there and train people on the standards for certification, let alone expect them to understand all the sections in that book. We also asked them to read the working papers series that were published by the National Council on Interpreting in Health Care. So it was really not a lot of preparation.

We also *recommended* that they take forty hours of basic interpreter training, such as *Bridging the Gap* or *Connecting Worlds*. And you’ll notice that I underline the word recommended. It wasn’t as though we said the pre-requisite for participating is that you have to have had this training. This is what they needed to have in order to prepare for the test: the *Bridging the Gap Interpreter’s Handbook*, *The Interpreter’s Edge* by Holly Mikkelson, the *Human Anatomy Coloring Book*, *Medical Terminology: A Short Course*, *Medical Standards of Practice*, and some medical dictionaries.

These were the lessons we learned regarding the test modules: We learned that the written test had its own value. The arguments for it are that the written modules measure basic knowledge. We asked people questions, and they could check-mark with pencil and paper. That could tell us their competency as far as entry level. It was noted in the report on national testing published by the National Council and written by Cindy Roat and Maria Paz Avery that the candidates who fail to demonstrate this basic knowledge should not be certified.

Now the argument against the written module is that, of course, written tests have no value in predicting the success of the overall language conversion. Written tests, as you know, cannot really determine how a person can perform and demonstrate her skills when she interprets. So again that was the lesson learned.

It was important of course to have language conversion, because that was a good predictor of how this person is going to do in the role play. And the role plays themselves were very, very good, as noted in the report. They were very authentic modes of assessment. But of course they were very costly to implement.

Test administration: Between Massachusetts and California, there were a lot of variations in the test administration. There were discrepancies and inconsistencies in that on the written test in Massachusetts, there was no time limit for modules number one and number two, the written tests covering standards and medical terminology. Yet, in California, because of miscommunication, somehow, we gave the participants a one-hour limit, which meant that most of the candidates could only complete module number one, which was the medical terminology. They were not able to complete module number two, which had to do with the standards of practice. And of course, again, I would like to repeat that at that time, standards of practice in California were barely known. We had just published them, and I doubt that a lot of the candidates in California knew much about the standards of practice that existed in Massachusetts.

The logistics for administering the language conversion were affected by opposite conditions. In Massachusetts, the test candidates came into a language lab and so the language conversion was proctored by administrators who controlled the process. In California, we did not have a language lab. The candidates came in. They had two tape recorders. On one tape, they listened to the prompts and then they answered with the other tape recorder. So they were the people who monitored their own answering, listening and all of that. So again, that was not quite consistent.

There were challenges in the administration of role plays. Again, there were too many discrepancies. The administrators were not consistent, allowing different pauses, for example. Inter-coder reliability was another issue. I believe that this was published in the final report. There was a very high variability in inter-coder agreement, and I think the possible contributing factors probably included failure to understand the coding system, not enough training, not enough detail in training, not enough explanation, lack of familiarity. All of these are factors that contribute to the high variability.

Just to put a human face on this whole process, I interviewed the raters myself. And this is what they said: "I think the pool of raters was very good, but because we had many different backgrounds, I think that some people were not as skillful as others in rating tests." "Some teachers approached the rating in a different way because they are more used to being consistent." Another comment here: "They told us how to rate the test, but they were not really explaining to us the rating process. I remember we had disagreements. There was a lot of disagreement regarding language equivalency or cultural issues, and so it was very ambiguous for most of the raters."

The implementation process suffered from lack of time and of course, lack of resources. Most of it was done by a lot of volunteer time, so that impacted the quality of the process. With regard to the testing instrument of course, we needed to be able to think of how to transfer the test over to other language pairs. I think we need to have a more stringent screening process, separating coders from administrators at the time of their testing. The people who administered were the same as the people who coded. The biggest challenge was to convey to coders what represents correct interpretation for concepts that didn't have equivalency. This was one of the biggest challenges for the coders. We needed more funding. We needed people who would know how to do coding. For the participants, this is what I would like to emphasize: more reliable

measures of participants' characteristics. To rely on participants to just declare how many training hours they've had is not an appropriate measure of the quality of training received. Nor can we rely on how many hours they've worked as interpreters to guarantee that they are competent either.

What has this joint project taught CHIA? To quote Don Schinske, our executive director, "CHIA does not have the organizational capacity to develop a certification test, set training standards, and promote certification for adoption by the state or, independently of the state, for dozens of health plans, medical groups, five hundred plus hospitals or eight hundred plus clinics." Consequently, our position is that CHIA does not endorse any certification process, but recognizes the major significance of the work done by others. We are willing and open to participate with all key stakeholders at either the regional or national level to explore the possibility of national certification.

Although CHIA, as an organization, does not have the resources to start a process for certification, we have identified some of the needs that must precede certification. We identified that it was important to have language proficiency assessment as a prerequisite for the test and for the training. People have to have standards of training, and then clear definitions for specific areas of training and knowledge. Then after the training, there must be some sort of post-training to see if the students are ready to interpret before they can even prepare themselves to take the test.

Because CHIA was not ready to undertake the certification process, the CHIA Standards and Certification committee disbanded and created the Education committee to start training and educating people. CHIA also proceeded to create an online registry of interpreters. It is like a directory with skill summary sections that allow the interpreters to go online and enter their contact information, and all information relevant to their work history and experience. At this point, there are no prerequisites -- such as years of experience or degrees -- to join the registry. However, it serves as an informed marketplace for healthcare interpreters who can compare their own skills with those listed by others.

Q: Building on this registry idea, is that currently up and alive?

A: Yes. Before I left, Don promised me that it will be online. But only members can access this at this point in time. So the requirement right now is that you have to be a current CHIA member in order to access that registry.

Maria-Paz Avery: Being that I helped developed the registry, I'm giving a presentation on this at the UMTIA conference on Saturday. And one of the things that Don empowered me to say, so I hope it's ok if I jump in here and say it, is that CHIA developed this with money from the California Endowment. CHIA is very interested, once it gets it up and running and has worked out all the bugs, to share both the technology, the program, all of it with any organization that is interested. He suggests that you might want to wait a couple of months so that CHIA incurs the cost of working out the bugs.

Q: So does that mean that health care providers that want to access the list would also then need to become members? Are there organizational members or just individual members? Do you have to be a member to search for an interpreter?

A: Yes, at this point in time. We have members who are organizations. We have managers of language services. We have interpreters. We have agencies. We have free-lance interpreters. So if you are a member you can access the registry, to search, and to look for the people whom you want to hire. In creating the registry we were hoping for people to realize that this is where the field is at this point in time. The message was: "We cannot guarantee quality as much as you want, but if you looked down this list and into this registry, you will see that there are interpreters who have more experience than others or more training than others, and that's a good point to start with".

Q: I just wanted to ask a little bit about the rater training. Correct me if I'm wrong, but is it safe to say that one of the lessons that you learned from this process was, based on some of the raters' comments, that the rater training was not for enough time? I've read some of the comments here, "Not enough, perhaps the selection process wasn't the best." I believe it says twelve hours of training. I'm just wondering how it was established that twelve hours would be sufficient? Or was it just kind of estimated based on what material needed to be covered?

A--Maria-Paz: It was based on the fact that we had very little funding. What would the funding support? Obviously, we knew from the very beginning that that was not going to be enough. Definitely for California. It was even less than Massachusetts, because I went there for two days and that was all we could do. And then I left, and came back to Massachusetts. And in Massachusetts, we continued beyond the two days because we could do it. And that's one of the lessons I'll talk about.

Q: So can I just ask quickly then how many hours training you did have in Massachusetts?

A--Maria-Paz: Well I don't know exactly, because we had the two days of training. And in addition to that, we continued to meet over the course of time that we were rating, just talking about issues and so on. In Massachusetts, we had also been spending a lot more time across the state with many interpreters, talking about our standards of practice, talking about what it meant, etc. So it was a very different context.

A--Elizabeth: Right. I think what happened in California was that the state was so big; the volunteers came from all over the place. We didn't have enough time for them. They were not very familiar with the MMIA standards of practice. We were not able to be there personally for them to kind of guide them through the process.

Maria-Paz: Somebody asked about the registry. MMIA has had a national and now an international registry for several years. At first it was just name of interpreter, address, phone number and email. And we changed it about a year ago to really relay the qualifications, educational experience, interpreter training, affiliation with other associations, and a few more items. And now, about three months ago, we created the possibility to upload the interpreter's resume onto that directory. So, I'd be happy to share what it looks like. But to the issue of being open or closed: many associations have an open registry where anybody can look into it. And what happens sometimes is

that there are security issues with regards to the information that's on the website, and also scams that occur that specifically target interpreters. So we took it off last August -- July or August -- and now only corporate members can access that directory. So being public or not is an issue. Some organizations have chosen, for liability purposes, to take information off the public domain. So, I'd be happy to share the information of some of our lessons learned.

Q: Was there a scoring guide when you did your rater training?

A--Elizabeth: I am sure there was a scoring guide.

A--Maria-Paz: We actually developed the rubrics for certain areas. So that was the scoring guide.

Q: So it was more primary trait, like you were looking at different areas instead of just underlined items?

A: Yes. Well, it depended on which part of the test it was.

Q: The other thing is did you have parallel versions? So were you dealing with different versions of the test?

A: No, it was exactly the same.

Q: Were the raters scoring one person and then that was calibrated so that they all came up with the same rating?

A--Maria-Paz: We were not that sophisticated. Again, you know we were doing this on zero dollars basically.