

## **A Proprietary Certification Program for Telephone Interpreters: Development and Insights**

Frances A. Butler, Ph.D., Language Testing Specialist  
Consultant to *NetworkOmni Multilingual Communications*

Good morning everybody. I'm very pleased to be here, and I wanted to thank Bruce for inviting me and asking me to participate in this interesting and important discussion that is going to be taking place over the next two days. As you know, I'm an independent consultant in language testing, and what I'm going to be reporting on this morning is a project that I was involved in with NetworkOmni Multilingual Communications on the development of a proprietary certification program for telephone interpreters. I've added to the title of the presentation *Development and Insights*, and following the request of Veronica and Bruce, I'm going to move fairly quickly through the details of the program and spend a little bit more time on insights from the work, challenges that I see as a result of that work, and the current effort that we're here to address.

First, I plan to talk about the background and rationale for our approach and then steps followed in the initial program development, then ongoing reassessment and periodic updates, insights on the application of testing principles in a new assessment environment, that is working with interpreters, and then issues and challenges in the current effort. Some background and approach: we began by putting together an interdisciplinary team of experts to work with NetworkOmni staff. Details about the program, who the experts are and so forth, are covered in Cindy's report for the California Endowment. So, for more specific information, I refer you to that document. There is also, I believe, a copy of an article in your binder that was in the *ATA Chronicle* that is about early stages of the development of the work at NetworkOmni. It is under Sawyer, Butler, Turner and Stone. So those two articles will provide more specificity about the program that I'm talking about today. We had an interdisciplinary team of experts that was composed of language testers, psychometricians, people who had expertise in cultural competence as well as interpreter training. The approach that we took was building content from an empirical base. This was very important to us and important for helping to establish the validity of the process. I'll talk a little bit more about that empirical base in just a minute. But from the empirical base, decisions were made about the makeup of the program. The program is modular, and it has a multi-skills assessment. By modular, I mean there are different pieces of the program that focus on training and testing, monitoring and so forth. Within the various phases of the program or modules of the program, we have a multiple-skills assessment that focuses on linguistic skills as well as interpreting skills.

The empirical base for the content, to go back to that for just a second, was important because for us it reflected the types of industries that NetworkOmni was working with and also the ranges of interpreting scenarios within and across those industries. I'll talk in a moment about how that was important feeding into the training and testing phases of the program. The steps followed in the development of the program were the following. We began with a needs analysis, and by this I mean we began by reviewing the existing processes. We went into this effort not starting from scratch because NetworkOmni already had a training and testing program in place, but what they were interested in was having us come in and help evaluate the program,

refine it and develop it into a certification program for them, including an internal certification exam for their telephone interpreters.

So we reviewed the existing processes. We documented the client base in terms of the types of scenarios that interpreters were being faced with in day-to-day interpreting situations, and the documentation of these scenarios allowed us to articulate the types of skills that interpreters were being called upon to demonstrate, and also then what we needed to test. As part of the needs analysis, we reviewed the client calls across industries, and this was very critical for establishing these patterns of calls across industries so that we could make decisions for selection purposes for training and testing. And the analysis of these calls, as I mentioned before, led us to the identification of scenarios. And these scenarios then eventually fed into test specifications and the design for the training program and for the testing program. So once we had gathered this information, we drafted a framework for test development, and for the development of all the materials in fact. And the drafted framework led to the various stages. We laid out those stages—screening, training, testing and monitoring, and in some instances these stages became iterative, in a way similar to what Janet was just talking about with Language Line. If at certain phases the interpreters who were tested did not pass the test, there was an opportunity for them to cycle back and receive additional training, additional monitoring and have an opportunity to be tested again at that particular stage before moving forward.

We used the information that we had gathered from the empirical work to develop and refine the materials and the test. We did this, as I mentioned, through the development of test specifications and then actually drafting items and beginning the piloting process, the tryout process, and revisions, again iterative. Those of you who have worked extensively in test development know that, as Janet was saying a minute ago, this is a process that requires a lot of going back and refining, trying things out again to get things to be the way you'd like them to be.

So once we felt confident about the quality of the procedures, we conducted a validation study with clients. We went out to a number of the different clients from different industries. Through the piloting process we had recorded a number of interpreters actually taking the test. We played these recordings for clients and asked them a couple of questions. First of all, we asked, in terms of the content of the test and the types of interpreting tasks that interpreters were being asked to do, "If interpreters could handle the material on the test, would you feel confident that they could handle most of the types of calls and so forth within your company?" That was one type of information we got back from the clients.

Next, the performances that we showed clients were different ranges of proficiency in terms of the effectiveness of the interpreters who had taken the test. We wanted to get a sense from them of the minimal level of proficiency that they would consider acceptable, that they would feel confident that the interpreters could handle everything in a high quality way for their personnel. So that was the validation study that we carried out with clients.

One of the things that we felt good about was that with the validation study we did not have to go back and make a lot of changes. Most of the clients were very positively disposed towards what we had produced in terms of the tasks. I credit that with our having worked from an empirical base. Before we actually did the test

development, we had a solid feel for the kinds of situations in which interpreters were going to be interpreting. So going from that base meant that we saved a lot of time later when we could have made mistakes if we hadn't had that solid background to build on. Working from an empirical base, I think, is always very important in a test development situation. Once we did the validity study with the clients, we then moved on to implementation of the internal program.

How a program is implemented is really important. It's something for us to think about in the current situation. With the NetworkOmni project, we didn't just implement the program and have it stop there. An important feature of the NetworkOmni program is the ongoing reassessment and updates to the program. These steps are critical for maintaining the validity of the whole process. I think it's important to remember that validity isn't a one-shot deal, but that it's an ongoing process that should continue for the life of any program. It's a part of keeping what you do current. So, on a periodic basis, NetworkOmni takes into account clients' changing needs. That could mean new clients with new needs coming in and being part of the program. It could also mean that there are internal changes going on in client companies, and that they may have different kinds of needs in terms of the types of interpreters that they need or the kinds of calls being handled and so forth.

Another element to think about is new approaches to training and testing. As we become more sophisticated in the techniques that we're able to use in testing and training, we want to be sure that those new techniques are implemented in an ongoing program. Yesterday someone mentioned the fact that changing technology has a lot to do with how we're able to deliver testing and training. So changing technology is an important piece of this too. And then as all these different changes are happening, it might mean the need to bring in additional outside expertise to help work on these various issues. So I think the main point here is that constantly monitoring the quality of your program means that you need to stay in close touch with your clients or with whoever the program is serving. As changes are going on, on a day-to-day basis, you need to figure out a mechanism for monitoring changes and getting the important ones implemented in whatever the program happens to be.

So now what I want to do is move on to the insights that I've had as a result of being able to work in this new testing environment. As a language tester and through the years I worked at UCLA I had an opportunity to work in many different language testing environments in this country and overseas, but until I came to NetworkOmni, I had never worked with interpreters. For me, it was an amazing opportunity and a challenge. It was extremely interesting in so many ways. One of the first things that I became excited to learn about was the complexities of interpreting beyond language proficiency. Clearly language proficiency in whatever the two languages are is a very important part of interpreting, and language proficiency must be assessed. But clearly, and this has come up in discussions already before now, being able to actually tap into interpreting ability beyond language testing is an interesting challenge. And how you define interpretation skills in the various milieus in which interpreters are functioning is part of what ultimately goes into the test specifications for testing interpreters.

A second issue for me was the special context of telephone interpreting. From one interpreting environment to another, you're going to have different situations that have to be addressed in different ways in terms of evaluation. Working on the NetworkOmni project really reinforced for me the need for an empirical base. I feel very

strongly that an empirical base is critical for helping to establish the validity programs and assessments within those programs. And there may be instances where you need to be able to show why you're testing in a certain way. The empirical information that you have provides the data that you can show the outside world, in terms of why the approach you're taking is the approach that you are taking. Also it will help you explain why you aren't doing something else. Another thing that I learned was the need for client education. It became very clear in a lot of instances that clients did not know how to work with interpreters. And so training clients to use interpreters effectively is very important, and I think actually somebody else mentioned that yesterday. I don't remember who it was.

Finally the issues and challenges I see in the current effort. One is the need for an articulation of what certification means in this context. We really need to be able to spell out what it is we're talking about when we say a person is certified. We had a little bit of discussion of that yesterday, and I think that in one of the smaller groups yesterday afternoon there was further discussion of that. But we need to be able to say what certification means and also what it doesn't mean.

Then, there's a need for clear specification of the content domains and sub-domains. I know there's been some discussion of whether we're talking about general community interpreting, or whether we're going to be talking about medical/health care interpreting. Whatever the decision is, in terms of the national certification, the content domain and the sub-domains have to be very clearly specified. And this is going to, I think, help feed into the articulation of what the certification means. We were talking a little bit in the testing group yesterday afternoon about this notion. Let's say we decided to focus on health care and medical interpreting. That would be a narrower focus than community interpreting, but still a very broad focus. And there is the question of how you make decisions for training and assessment. Everything cannot be covered or tested, so selection priorities must be set. Setting those priorities is one of the things that we need to do.

In closing, I just want to reiterate something I said a minute ago, and that is we always need to remember that validation is a process and that it's critically important to document every stage of the process. The meeting that we're attending here today is a step in a process. And I would say that depending on where the focus is in terms of community or health care interpreting, if it turns out that the standards that have already been developed are part of this process, then it's important to document how those standards were developed. In addition, there should be outside validation of those standards if they are going to be critical in terms of informing the decisions that are made.

---

Q: I just wanted to ask about when you say multiple skills assessment. What did you mean exactly by that?

A: I meant that in the scenarios, in the testing, we were tapping both linguistic elements and interpreting skills. I didn't go into a lot of detail, in terms of diagram and sort of the flow of how the system is set up, but first there's a screening process, where language proficiency alone is initially tested for the two languages. Then as we move into different

stages of training, testing and monitoring, at the later stages, language proficiency or linguistic skills continues to be part of the evaluation process at every level.

Q: You talked about a customer satisfaction kind of study, and I understand that. But your other study, where you asked clients what would be the minimal level of proficiency they would be comfortable with, and you played I guess different levels of interpretation, could I get some more details on that?

A: I'm not sure I can. Jean is here. I should say Jean was part of the team. Do you want to respond?

A (Jean Turner): Sure. As I recall, and I believe I recall accurately, there were three major client groups that were identified, and then we met with three clients within each one of those client groups. And there were two of us. Sometimes it was David Sawyer and I, or Frances and David, or some combination of the three of us who visited each one of these clients. And typically the people that we met with included the supervisor or the customer service representatives for the client, and a number of people who were customer service representatives who were bilingual. Because there was some question of how do you know whether the interpreter's doing a good job unless you know the languages? And if you need an interpreter, you don't know the languages. So how do you figure that out? So that was why we specifically requested there would be people that would know the languages as well, so that this process of listening to calls they could respond to something other than just whether their voices were nice.

Q: Were they interpreters or just some kind of bilingual...?

A (Jean): Two of the groups that I participated in, client groups, there were one or two or sometimes three people who actually served as interpreters themselves for the companies.

Q: So did you play failing performances?

A (Jean): I wouldn't say that they were failing, but that they were poor, weak.

A: And I'm not sure. I may have misspoken when I said we asked them to identify a level. We didn't use the term level. I guess maybe in our minds, we were trying to see it. The tapes that we played showed a range of ability or effectiveness or whatever terminology you want to use.

Q: But were any of them failing?

A: No.

A (Jean): We didn't think it would be productive or good P.R. to write a failing list because in fact people who failed at the initial stages of the program are not working for the company because they demonstrated that even with training and support, they couldn't handle the work. So, no.

Q: It just occurred to me. I'm sorry I didn't ask. Is this a medical certification program or just general or all different types?

A: It's a general certification program to begin with, and there's a module that focuses on medical.

Q: So, there's a separate test for medical and a separate one for another module?

A: It was constructed that way so that over time, as the client base might change, different areas could be added. But the initial approach is that we wanted to tap first basic interpreting skills that could cut across the different industries. And actually, I don't think I mentioned earlier, but as part of the empirical work, when we were looking at the types of calls that were coming in from across industries, we were looking at four patterns or types of calls and types of scenarios and so forth that would address the general interpreting skills.

Q: I was just going to ask you what domains or competencies actually were tested or are tested for certification? You mentioned language proficiency, but in terms of interpreting, do you test auxiliary interpreting skills? I'm just trying to understand what it is that you test.

A: General interpreting skills, abilities to transfer from one language to the other.

Q: So, consecutive interpreting skills?

A: Yes.

Q: Any other skills? Like sight translation?

A: No translation, no.

Q: So you said it's a basic skills test, and you have modules. But you're testing company bilingual employees? So different positions, different jobs have different modules that you test?

A: Well, there's the potential there for developing modules. But a lot of the work is customer service across a range of industries, so that's the general test that the program covers, the training and testing. But then it also offers the opportunity for more specific modules, and health care was one of the areas where a different focus was needed. So in addition to the general customer service type of material, health care was added as a separate module. And it would be possible to add other modules if the type of work became more specific than what had already been identified from the general interpreting data.