

 *Exploring the Validity of the Transsexual Voice Questionnaire for Male-to-Female Transsexuals*

 *Explorer la validité du « Transsexual Voice Questionnaire » appliqué aux transsexuels d'homme à femme*

KEY WORDS

TRANSSEXUAL

VOICE

QUESTIONNAIRE

Shelagh M. Davies
Judith R. Johnston

Abstract

Speech and voice feminization for transsexual women is a relatively new field within the scope of speech-language pathology and research in this area is sparse and narrowly focused. The current study investigates the content validity of a newly developed tool, the Transsexual Voice Questionnaire, Male to Female, TVQ^{MTF}, (Dacakis, Davies, Oates, Douglas, & Johnston 2013). Five transsexual women were interviewed and the content of the interviews was compared with the content of the TVQ^{MTF} items. The themes most frequently expressed in the interviews were the themes addressed in the TVQ^{MTF}. Similarly the items on the questionnaire that were rated as particularly problematic concerned the themes that were most frequently raised in the interviews. Although the sample size was small these initial findings indicate that the TVQ^{MTF} may provide a valid picture of transsexual women's real life experiences with voice and is worthy of further study.

Abrégé

La féminisation de la parole et de la voix pour les femmes transsexuelles est un domaine relativement nouveau pour la discipline de l'orthophonie, et la recherche dans ce domaine est rare et étroitement ciblée. Cette étude explore la validité du contenu d'un outil nouvellement développé, le *Transsexual Voice Questionnaire*, d'homme à femme, TVQ^{MTF}, (Dacakis et al. 2013). Cinq femmes transsexuelles ont été interviewées et le contenu de ces entrevues a été comparé au contenu des éléments du TVQ^{MTF}. Les thèmes exprimés le plus fréquemment dans les entrevues étaient ceux qui étaient d'abord soulevés dans le TVQ^{MTF}. De façon semblable, les éléments du questionnaire qui avaient été jugés particulièrement problématiques étaient les mêmes que ceux qui avaient été fréquemment rapportés dans les entrevues. Quoique la taille des échantillons était petite, ces constatations initiales indiquent que le TVQ^{MTF} pourrait offrir un portrait réaliste des expériences de vie de femmes transsexuelles pour ce qui est de la voix et mérite une étude plus approfondie.

Shelagh M. Davies, M.Sc.,
Clinical Assistant Professor,
School of Audiology and
Speech Sciences
Faculty of Medicine
University of British Columbia
Vancouver, BC
CANADA

Judith R. Johnston, M. A., Ph.D.,
School of Audiology and
Speech Sciences
Faculty of Medicine
Vancouver, BC
CANADA

Speech and voice feminization for transsexual women is a relatively new field within the scope of speech-language pathology. In their *Standards of Care* the World Professional Association for Transgender Health (WPATH) defines 'transsexual' as "Adjective (often applied by the medical profession) to describe individuals who seek to change or who have changed their primary and/or secondary sex characteristics through feminizing or masculinizing medical interventions (hormones and/or surgery), typically accompanied by a permanent change in gender role." (WPATH, 2013, p. 103). For transsexual women the sense of 'being born in the wrong body' may be acute, leading them to take steps to feminize aspects of communication such as speech and voice, to harmonize with their own sense of gender.

Prior Studies

In 2012 Oates reviewed the literature discussing communication changes among transsexual women and found that most papers were chiefly concerned with voice. An earlier study (Pasricha, Dacakis and Oates, 2008) noted that voice was also the aspect of communication with which the transsexual women themselves were mostly concerned. Early investigators into voice feminization in transsexual women looked at the average speaking fundamental frequencies (SFF), equating higher SFF with good intervention outcomes. Trends in the Gelfer and Schofield (2000) data indicated that transsexual women subjects perceived as female did have a higher SFF and higher upper limit SFF than subjects perceived as male.

Other researchers have expanded this line of investigation to include the effect of other acoustic and aerodynamic parameters on perceived speaker gender. Holmberg, Oates, Dacakis, and Grant (2010), for example, found that the use of low speech intensities and increased glottal airflow could contribute to a successful female voice. Studies by Günzburger (1993) and Carew, Dacakis & Oates (2007) largely agree with these findings. Carew et al (2007) and Hillenbrand and Clark (2009) suggest that a combined modification of both SFF and resonance tract formants may be most effective in feminizing the voice. Hancock and Garabedian (2012) come to a similar conclusion but also note that modifications of speech and voice other than SFF and resonance have not been widely studied. Recently Hancock, Colton, and Douglas (2014) suggest upward gliding and larger semitone range of intonations may contribute to feminizing the voice in transsexual women speakers.

Reports by McNeill, Wilson, Clark and Deakin (2008) and Owen and Hancock (2010) have expanded the discussion

beyond the link between objective parameters of voice and perceptions of gender. McNeill et al (2008) suggested that simply achieving a higher SFF did not guarantee a transsexual client's happiness with her voice. Both these studies examined the relationship between a transsexual woman's voice and ratings of femininity, as perceived by both herself and her listeners. Both studies found that a higher self-rated degree of femininity corresponded with a higher listener-rated perception of vocal femininity.

This short literature review highlights the expanding areas of interest in the field of transgender communication. Davies, Papp, and Antoni (in press) give an extensive review of this literature.

Laboratory studies of feminine voices have identified some of the gender-related acoustic features of speech and have demonstrated that listeners can and do make judgments of gender based on vocal cues. Researchers have not yet tried to connect the acoustic and perceptual facts to the real life experiences of transsexual persons. This may in part reflect the absence of an appropriate measure of "real-life".

In 2006 an assessment tool, The Self Evaluation of Voice Questionnaire (TSEQ), was published (Davies and Goldberg, 2006a; 2006b). This questionnaire was modified from the Vocal Handicap Index (Jacobson, Johnson, Grywalski, Jacobson, & Benninger, 1997) to be trans-specific. It was designed specifically to measure a transsexual person's concerns with voice and to provide a broad picture of a transsexual individual's experience with voice and the ways that voice relates to sense of self and daily life. In the years since its publication, the TSEQ has been included in feminization protocols by increasing numbers of clinicians and researchers throughout North America and abroad. Although the widespread use of this questionnaire would suggest otherwise, little was known about its characteristics. In particular, the psychometric properties of the TSEQ had not yet been established.

Current Study

In 2008, researchers from La Trobe University in Melbourne and the University of British Columbia in Vancouver began a collaborative project that was initially intended to determine the psychometric properties of the TSEQ. As a first step, the researchers reviewed each item and made revisions that would improve clarity and relevance. These decisions were guided by consultations with transsexual clients and speech-language pathologists as well as the researchers' own clinical experience. The resulting revision of the TSEQ was extensive: 24 of the

30 items in the TSEQ were replaced or significantly modified. (See Dacakis, Davies, Oates, Douglas, & Johnston, 2013 for a detailed account of the revision process.) The item review process had essentially created a new questionnaire. To avoid future confusion this revised version of the TSEQ was renamed the Transsexual Voice Questionnaire, Male to Female (TVQ^{MTF}), and the researchers shifted their investigation of psychometric properties from the TSEQ to the TVQ^{MTF}. Like its predecessor the TVQ^{MTF} has 30 items, each asking the user to rate the frequency of an experience or the strength of agreement with a statement, using a 4 point Likert Scale.

Two studies have now been conducted to provide an initial look at the psychometric properties of the TVQ^{MTF}. The first study investigated reliability (Dacakis et al, 2013). Chronbach's alpha, item-total correlations and intraclass correlations were used to determine internal consistency and test-retest consistency. The results of these analyses all indicated that the TVQ^{MTF} was a reliable assessment instrument (Dacakis et al, 2013).

We are reporting the second study here - an investigation of the content validity of the TVQ^{MTF}. We used data from interviews with transsexual women to address the question: To what degree does the Transsexual Voice Questionnaire, Male to Female (TVQ^{MTF}) provide a valid picture of the concerns of transsexual women about their voices?

Approaches to Validity

Attempts to define validity reach back more than a century. In spite of much effort, validity remains "possibly the most fundamental and the most elusive of all assessment concepts" (Newton, 2013). Theorists writing about validity run the gamut from a positivism that celebrates universals and the science of discovery to a postpositivism that no longer looks for truth but accepts the local and constructed nature of knowledge (Lather, 1993). These philosophical differences are reflected in the types and subtypes of validity that are proposed. Cook and Campbell's (1979) types of validity, i.e. statistical conclusion validity, external validity, and internal validity are clearly different than those discussed by Lather (1993) from her postpositivistic position, i.e. ironic validity, paralogical validity, and voluptuous validity.

A paper published in 1991 by S. Messick of the Educational Testing Service borrows from each end of the continuum – traditional categories of validity but interpretations that are situated in particular moments of

practice. "Validity is an integrated evaluative judgment of the *degree* to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores and other modes of assessment." (Messick, 1991, p.5). Messick argues that validity is not so much a property of a test or other assessment tool as it is a complex decision to be made about an entire body of evidence and theory. The value of any single concept or piece of evidence will depend on the nature and adequacy of past work and the anticipated consequence of the validity decision at a specific time and place.

Given these different views on the fundamental nature of validity, it is not surprising to find different views on content validity as well. In some instances content validity is treated as subordinate to construct validity, in other instances it stands alone. Content validity can be viewed as a characteristic that an assessment tool does or does not have, or as a piece of information that varies in importance from one decision to the next depending on the larger context of data, theory, values, and circumstance. Despite such differences, theorists who write from within the more traditional frameworks are likely to include content validity in their taxonomies and there is considerable agreement about what it entails and how it can be measured.

A working definition of content validity is provided by Haynes, Richard and Kubany, (1995), who state that "content validity is the degree to which elements of an assessment instrument are relevant to, and representative of, the targeted construct" (p.239). A 'targeted construct' is whatever the assessment tool is meant to measure. The TVQ^{MTF} is intended to identify and measure the experiences and concerns of transsexual women about voice and voice change. There are a number of different ways that content validity can be determined, but two in particular seem to fit the size and exploratory nature of our study: criteria-relationships and concurrency.

The first approach assesses validity by establishing a standard against which the TVQ^{MTF} can be evaluated. In our case the standard was the interview content. We examined various aspects of the interview data and determined the degree to which the TVQ^{MTF} agreed with our standard. The second approach takes a slightly different perspective on the data and asks whether or not individual responders performed similarly on the two measures, i.e. did a person who talked at some length about the social consequences of voice also rate social experiences as challenging on the TVQ^{MTF}.

METHOD

This research project was approved by the Behavioural Research Ethics Board at the University of British Columbia, Vancouver, Canada.

Participants

Five women who self-identified as transsexual were the primary participants in this study. They were recruited from the community through posted notices in clinical offices and short talks by the first author at a transgender support group. The five women were a subset of a larger pool of volunteers (6 women, 5 men) and were selected with the following concerns in mind. First, persons in the 'transgender' category group cross or transcend culturally defined categories of gender in diverse ways, including differences in the degree of disparity between their gender identity and the sex they were assigned at birth (WPATH, 2013). In order to make the findings in this initial project as clear as possible, we decided to limit the study to those who identified as transsexual rather than those who fell within the more heterogeneous category of transgender. Secondly, as the study progressed it became evident that the experiences with, and concerns about voice were quite different for transsexual women than they were for transsexual men. The differences stemmed in part from differences in the outcomes of hormone therapy. The androgens frequently prescribed to transsexual men may cause a profound and desired decrease in SFF and SFF range without further intervention. (Damrose, 2009). The hormones given to transsexual women have no analogous effect. To remove this source of heterogeneity in our sample we decided to limit our study to transsexual women. One woman chose to discuss topics unrelated to voice and her experiences with voice, so her data were also excluded from further analysis.

Our final participants were 5 transsexual persons, all Male to Female (MtF). Their ages ranged from 44 to 64 with a mean age of 54.8. All participants were fluent in English and none were communicatively impaired. The length of time presenting as female ranged from 1 year in trans-friendly environments to 7 years, 4 months in all situations. All participants were using feminizing hormones and had been using them for 5 months to 7 years, 4 months. No participants had undergone pitch elevating phonosurgery, 2 out of the 5 had participated in voice therapy and only one had undergone gender reassignment surgery.

Procedures

Each participant was seen once for approximately

one hour. The session began with a 30 minute interview exploring their experiences and concerns about their voices. The participants then filled in the TVQ^(MTF) after which they were asked to comment on the TVQ^(MTF) and add any further concerns about voice which had not been elicited in the interview.

A graduate student who had been trained in ethnographic methodology conducted and audio-taped the interviews, beginning with open-ended questions, e.g. "Can you tell me about your voice?" When responses to the very general questions were no longer forthcoming, the interviewer asked a few more specific questions to see if the participant had further ideas to share, e.g. "Are there any situations in which you're more aware of your voice?" Recordings were transcribed by a speech-language pathology graduate student. The transcripts were then reviewed for accuracy by the first author as she listened to the audio recordings. (Heilmann et al, 2008). This approach to transcript reliability has been developed and evaluated extensively by Jon Miller and his colleagues at the Language Sample Analysis Laboratory at the Weismann Centre, University of Wisconsin, Madison (Heilmann, et al, 2008). It is more economical than the traditional procedure of comparing two independently prepared transcripts of the same session and yet yields high levels of accuracy.

Discourse and Content Analysis

In order to reach our research goals we needed a way to describe the interviews and the TVQ^(MTF) data in the same terms. We developed an analysis scheme by looking closely at the questionnaire and identifying the recurring themes that seemed to underlie the questions. We found six such themes:

1. Effect of voice on ease of social interaction
2. Effect of voice on emotions
3. Relationship between voice and gender identity
4. Effort and concentration required to produce voice
5. Physical aspects of voice production
6. Pitch

This scheme was used to code the content of both the TVQ^(MTF) and the interviews. Each of the TVQ^(MTF) items and each idea expressed in an interview was tagged as belonging to one or another of the six content themes. Items that could not be assigned to a theme were noted as such. The two authors and a senior colleague made the decisions about content and will be referred to as "coders." All three were experienced speech-language pathologists

with considerable expertise in discourse analysis, voice assessment, and/or the analysis of spontaneous conversational samples. All final codes were consensus decisions; disagreements were resolved by discussion.

Coding of the TVQ^(MTF) required decisions about the general topic area that each questionnaire item represented. Table 1 summarizes these relationships. As can be seen, 29 of the 30 items could be classified in one or another of the six content areas of our coding scheme, with the number of items per theme varying from 3 to 7. This correspondence is not surprising since the themes were initially drawn from the TVQ^(MTF).

The only item remaining outside one of the 6 content areas was #23: “My voice restricts the sort of work I do”. This item turned out to be a loner. There were no other TVQ^(MTF) items on the topic of employment and it was not raised by anyone being interviewed. Since our project was not designed for item-by-item analysis, we are unable to comment further.

Coding of the interview transcripts was a more complex task that required decisions about three aspects of the discourse: turn boundaries; presence or absence of substantive ideas; and the meaning category of each idea identified. Basically, the coders had to decide what, if anything, the interviewee was talking about each time she

Table 1: The items that fall into each of the content themes derived from the TVQ^(MTF).

Theme	Item#	TVQMtF Items	Total items in theme
#1: Effect of voice on social interaction	7	I avoid using the phone because of my voice	7
	8	I’m tense when talking with others because of my voice	
	12	I feel uncomfortable talking to friends, neighbours and relatives because of my voice	
	13	I avoid speaking in public because of my voice	
	17	My voice difficulties restrict my social life	
	25	I am less outgoing because of my voice	
	26	I feel self-conscious about how strangers perceive my voice	
#2: Effect of voice on emotion	2	I feel anxious when I know I have to use my voice	6
	14	My voice sounds artificial	
	16	I feel frustrated with trying to change my voice	
	24	I feel my voice does not reflect the true me	
	28	It distresses me when I am perceived as a man because of my voice	
	30	I feel discriminated against because of my voice	

#3: Relationship between voice and gender identity	3	My voice makes me feel less feminine than I would like	5
	6	My voice gets in the way of me living as a woman	
	10	My voice makes it hard for me to be identified as a woman	
	19	When I laugh I sound like a man	
	20	My voice doesn't match my physical appearance	
#4: Effort, concentration needed to produce voice			3
	15	I have to concentrate to make my voice sound the way I want it to sound	
	18	When I'm not paying attention my pitch goes down	
	21	I use a great deal of effort to produce my voice	
#5: Physical aspects of voice production			4
	1	People have difficulty hearing me in a noisy room	
	9	My voice gets croaky, hoarse or husky when I try to speak in a female voice	
	22	My voice gets tired quickly	
	27	My voice "gives out" in the middle of speaking	
#6: Pitch			4
	4	The pitch of my speaking voice is too low	
	5	The pitch of my voice is unreliable	
	11	When I speak the pitch of my voice does not vary enough	
	29	The pitch of my speaking voice is restricted	

took a conversational turn. For the purposes of this project an "idea" was defined as a topic consistent utterance or set of utterances. A given idea ended with the introduction of a new theme-topic. To be counted as a distinct "idea" a passage needed to focus on a topic that could be distinguished from other topics in the surrounding text. This notion of a general thematic "idea" is similar to what has elsewhere been called a "discourse topic" (Brown & Yule, 1983) as opposed to a propositional topic. An example of the analysis is provided in the Appendix. Note that our definition does not rule out redundancy; a given idea can persist throughout an interview.

The three coders read the transcripts and identified and classified each idea about voice that was expressed in the interviews. They also noted when the interviewee talked about an aspect of voice for which there was no coding category or introduced topics that did not concern voice.

Although there were only five participants, the interview database was quite substantial. The coders reviewed some 780 utterances spoken by the transsexual women, divided these utterances into 196 turns, and then within those turns identified and classified 316 ideas, 240 of them were explicitly about voice.

Results

Although the interviews were focused on voice, preliminary analyses revealed that, on average, 24% of the ideas expressed did not specifically concern voice, and while these topics were appropriate in context, they could not address our research questions. Among other things the interviewees talked about aspects of gender presentation other than voice, and language forms and patterns often identified with female and male speakers (Tannen, 1990). They also asked for clarification, or in some other way managed the conversation. In the analyses that follow, we will only use the 240 ideas expressed about voice.

Tests of Criteria-Relationship

Our first specific research question was, “Do the TVQ^(MTF) items cover the same concerns about voice as are spontaneously raised by transsexual women?”

We began with a general comparison of the thematic content of the TVQ^(MTF) and the thematic content of the interviews. Taking the interview data as the criteria we found that virtually all of the TVQ items (29 out of 30) do address the concerns and experiences voiced by the interviewees. This is one indication of questionnaire validity. However, it was also true that 24% (N=58) of the voice related ideas

from the interviews did not fit into the TVQ^(MTF)-based thematic categories.

The most frequently mentioned area of concern not found in the TVQ^(MTF) was voice change. Comments and questions about the likelihood of voice change and the role of therapy in this process comprise 10.7% of the ideas expressed. This broad category is shown in Table 2, column 7.

The remaining 13.2% of the voice related interview content was comprised of an assortment of topics, briefly discussed, often by only one person: surgery, non-speech vocalization, such as laughing or coughing, speech and voice variation depending on conversational partner, and the holistic nature of the feminization process. These topics are grouped together in column 8, not because they share content, but merely to account for the remaining percentage points.

Thus far the content comparisons had showed that while virtually all of the TVQ^(MTF) items dealt with ideas that had been spontaneously raised by the transsexual women, the questionnaire did not address the full array of their concerns. However, looking further at the six interview topics that did not fit within the thematic categories of the TVQ^(MTF), they seemed to be topics that were inherently low frequency. We tested the relationship between frequency

Table 2. Individual and mean percentages of voice-related ideas expressed in the interviews for each content theme in the TVQ^(MTF) (a).

ID	TVQ ^(MTF) Content Themes							
	#1 Social Interact	#2 Emotions	#3 Gender Identity	#4 Effort	#5 Physical Aspects	#6 Pitch	#7 Change Process	#8 Other Voice
1.	9.0	4.5	19.4	14.9	3.0	22.4	7.9	19
2.	4.6	6.8	27.0	6.8	15.9	22.7	0.0	5
3.	4.6	9.1	27	13.6	3.0	9.1	14.3	19
4.	6.3	18.8	31.3	12.5	0.0	12.5	12.5	6.3
5.	3.3	3.3	40	3.3	3.3	20	10	16.7
Mean	5.6	8.5	29.	10.2	5.1	17	10.7	13.2
(sd)	(2.2)	(6.2)	(7.5)	(5)	(6.2)	(6.2)	(2.6)	(7)

a. All values in this table are percentages. The % symbols were omitted to improve readability.

and inclusion in the questionnaire, using Fisher's Exact Probability Test for a 2 X 2 cross tabulation, i.e. themes that occurred in the interviews with High and Low frequency crossed with interview topics that were In or Out of the questionnaire. The relationship proved significant, $p=.040$. The TVQ^(MTF) items primarily dealt with the ideas that occurred more frequently in the interviews. There is a practical limitation on the number of items that can be included in a questionnaire. Given that the whole array of topics could not be included, the fact that the TVQ^(MTF) tends to include those topics that are more likely to be raised by the transsexual women themselves is good evidence of representativeness.

Our last criteria-related analysis looked further at the representativeness of the TVQ^(MTF), using the data that related to our initial six themes. In the cross-tab analysis it was largely these themes that had been grouped together as being High Frequency. We now treated each theme individually and calculated the proportional frequency of mention for each theme in each of the five interviews. Table 2 provides both the individual and the group data. It is clear from this table, and from Table 1, that neither the questionnaire items nor the ideas expressed in the interviews were distributed evenly across the six themes. Some ideas were expressed relatively more often than others and there were more items for some themes than others. We calculated a Spearman rank order correlation between the distribution of ideas in the interviews and items in the questionnaire. A correlation coefficient of $-.25$ indicated that there was no reliable relationship between these distributions. The TVQ^(MTF) items dealt with content themes that were also voiced by our transsexual interviewees, but not proportionally. Ideas that occurred more frequently in the interviews were not necessarily the ideas that were explored with more questionnaire items.

To aid in the interpretation of this finding, we looked again at the data in Table 2. The standard deviations for some of the themes indicate similar degrees of interest across the five women. Other themes have relatively large standard deviations because one or two of the women show a much greater interest in the topic than do the rest of the women. A larger database would be needed to describe and interpret such differences. As a starting point however, we calculated Kendall's coefficient of concordance to see the degree to which the 5 participants agreed in their choice of themes to talk about, and at what length. The W coefficient was $.63$, significant at $p<0.01$. This value indicated a moderately strong level of agreement among the participants as to the relative importance of the various themes.

To summarize our criteria-related tests of validity, we found good evidence of both relevance and representativeness. Relevance is seen in the fact that 29 of 30 items in the questionnaire focused on concerns and experiences that had been raised by transsexual women in the interviews. Assuming that importance is reflected to some degree in frequency of mention, there was considerable agreement among the women as to the relative importance of the various themes. Representativeness is seen in the fact that the themes most likely to be included in the TVQ^(MTF) were the themes that had high frequencies of mention in the interviews.

Tests of Concurrency

We turn next to analyses designed to answer our second research question: "Do the rating responses of transsexual women on the TVQ^(MTF) indicate the same degree of relative concern for a topic as was indicated in their interviews?" This is essentially a test of concurrent validity. The women participated in an interview about their voice and then later on the same day completed the TVQ^(MTF). We look at the question, "To what degree do their responses in these two tasks agree?"

Each TVQ^(MTF) item requires the participant to respond by checking a number from 1 to 4. Higher values indicate greater levels of concern or higher frequencies of occurrence. Our research question was explored by looking to see whether the issues raised and discussed most often in the interviews were the same as the issues that had been checked with higher numbers on the TVQ^(MTF). For each of the TVQ^(MTF) themes we took the numbers checked on all of the items belonging to that theme by all of the participants, and calculated an average response score. These average scores represented the degree of participant concern about each of the content areas. For example, as can be seen in Table 1, there were three TVQ^(MTF) items (15, 18, 21) belonging to Theme 4, "effort and concentration required to produce voice". The average response score for this theme was 3.5. In contrast, the comparable mean score for the seven items belonging to Theme 1 "the effect of voice on the ease of social interaction" was only 2.5. We rank ordered the average response scores for each of the six TVQ^(MTF) content themes, and compared that rank order to the rank order based on the relative number of times a given theme-idea was expressed during the interviews. The Spearman rank order correlation coefficient for this comparison was $.77$, indicating a strong level of agreement. Due in large part to the small number of comparisons, this value did not reach statistical significance. Nevertheless, for any given theme, if the average response score on the TVQ^(MTF) was high, the

average occurrence of that theme in the interview data also tended to be high.

As a final test of validity, we used percent-of-responses-at-level-“4 (usually or always)” as another measure of the level of concern about each theme. Again combining the data from all of the items belonging to a theme and from all of the participants, we calculated the percent-of-responses-at-level-4 for each content theme. We compared the rank order of these percentages with the rank order of the themes as indicated by frequency of mention during the interview. The rank order correlation coefficient for these two variables was .89, statistically reliable at $p < .05$. This coefficient indicates a very strong agreement between the level of concern about a given content theme that was evident in the TVQ^(MTF) responses and the level of concern evident in the interviews. Interview topics that were raised and discussed more often were more often rated with “4”.

Discussion

Major Findings

This study was designed to determine whether the Transsexual Voice Questionnaire (Male to Female) provides a valid picture of transsexual women’s concerns and experiences with voice. We attempted to answer this question by comparing features of content, structure, and response in the TVQ^(MTF) with the views expressed by transsexual women themselves. Specifically we asked (a) if the content themes in the TVQ^(MTF) corresponded with the content of the interviews, (b) whether the proportional distribution of TVQ^(MTF) items across themes mirrored the distribution of ideas found in the interviews, and (c) if the themes rated as problematic on the TVQ^(MTF) were also the themes that were discussed more frequently and at greater length in the interviews. These are the major findings from these analyses:

- All of the content themes in the TVQ^(MTF) also occurred in the interviews.
- The TVQ^(MTF) content categories did not cover all of the ideas expressed by the interviewees, but the ideas that were covered tended to be high frequency, and by inference, important.
- Severity ratings on the TVQ^(MTF) were in accord with the likelihood and extent that a given theme would be raised and discussed in the interview.

Taken together, these findings provide empirical evidence of the validity of the TVQ^(MTF), or in the framework of Messick, they would contribute to an *integrative evaluative judgment* that the TVQ^(MTF) could be used in a given situation.

Research Implications

Using Scores

The one test that did not have a positive outcome was our detailed test of proportional representation. According to Haynes et al. (1995), similarities in proportional distributions would be one way to demonstrate representativeness and ultimately the content validity of the TVQ^(MTF). We found that the proportional distribution of *items* across themes did not match the proportional distribution of *ideas* across themes. The interviewees and the questionnaire differed as to which themes deserved the greatest attention. While proportionality would indeed contribute to the validation process, its absence is more difficult to interpret. In the present study it seems that the scope of some themes, e.g. ‘the effect of voice on social interactions’, is very wide and naturally invites a larger number of items than themes such as ‘pitch’. If this is true, an even distribution of items across themes would present a distorted view of voice concerns. Clinicians and researchers who wish to “score” the TVQ^(MTF) should keep in mind the uneven distribution of items across themes and either use theme weights when calculating total scores or avoid overall scores and instead use scores for sections defined by content themes .

Questions about scoring would be best answered by multivariate analyses but the small size of our database precluded the use of these approaches. With a larger database it would be possible to use procedures such as Discriminant Function Analysis and Factor Analysis to address questions about individual differences or about the dimensions or factors that underlie responses on the TVQ^(MTF). These factors may resemble the topic themes used in the present study, or they may point to some aspects of personality or circumstance that have broad influence such as optimism, sociability, or self-directedness. In either case, scoring systems could be developed from the dimensional data such that the total score represented each dimension equally or gave more weight to dimensions that reflected the client’s values and goals.

Future directions

The positive findings reported in both the Dacakis et al (2013) study of reliability, and the study of validity reported here invite further investigation of the new questionnaire. Given the small number of participants in the current study our findings can certainly be viewed as preliminary, but they also can be viewed as initial, i.e. as the first steps in the development of the body of evidence needed to guide assessment decisions.

We are still in the early stages of understanding gender markers in voice and communication and how to affect their change in transsexual people. The current body of literature is small, with small numbers of participants in the studies and lack of control groups. In her review of the literature Oates (2012) notes that the current evidence for voice therapy's effectiveness in this population is weak. She found that the data reported in 83% of the studies consisted of expert opinion or consensus - "the very lowest level on the evidence hierarchy. The remaining 17% of publications in this field provide only marginally stronger evidence" (Oates, 2012, p. 59). We need larger studies designed to provide higher levels of evidence to confirm, disconfirm or nuance the initial findings reported here.

Research Priorities

Findings of this research project have clear implications for clinical researchers wishing to start new lines of inquiry. As indicated in Table 2, "The relationship between voice and gender identity" was the theme most frequently raised in the interviews. Not only did it have the highest mean value for percentage of ideas expressed, 29%, but it was ranked in importance as number 1 or 2 by each of our five participants. As one participant put it, "The voice is how you express that [the gender transition] journey." This concern with voice was also expressed by the transsexual women that Dacakis et al. (2013) interviewed during the development of the TVQ^{MtF}. Yet in the larger clinical and research literature concerning gender transition, voice is rarely discussed. This suggests a critical need for input from speech-language pathologists and linguists about the role of voice in gender transition.

Theme #6, 'Pitch' was rated as the second highest in importance by the interview participants, and contained 17% of spontaneously raised content. This is not surprising. Pitch is perhaps the easiest perceptual variable to hear in a voice and there are established norms for average speaking pitch ranges for cis (genetic) men and cis women (Hillenbrand & Clark, 2009). However what is commonly understood as 'pitch' includes contributions both from the larynx (fundamental frequency and harmonics) and vocal tract (formant frequencies). It is perceptually difficult to distinguish between these two inputs and most probably the women in this study have the layperson's more general understanding of the term 'pitch'. The complex interplay of laryngeal and vocal tract contributions to pitch perception is an area that would welcome research and would have clinical applications.

Clinical Interpretations and Applications

As is true with any assessment tool, effective use of the TVQ^{MtF} increases with knowledge and experience. The

following points may be useful to remember while using this questionnaire. First, it is important to keep in mind that TVQ^{MtF} responses are Likert scale responses. As such they can convey only order, not unitized quantity. If a transsexual woman originally responds to a question by marking "4" and later responds with a "2", we know that she is now less concerned about that matter but we do not know how much less.

Knowledge of the psychosocial background of the individual completing the TVQ^{MtF} may also prove useful when interpreting the responses. This seems particularly true of the questions associated with Themes 1 and 2. Item 7, for example, reads "I avoid using the phone because of my voice." With today's technology, it is hard to imagine the circumstances that would lead to a "4" for that item. On the other hand, a "4" rating on item 13, "I avoid speaking in public because of my voice" is easy to imagine. Clinicians may wish to decide item by item which answers imply the more severe social problem.

In some cases it may even be that the absence of a problem is the problem. Consider two possibilities for rating item #17 "My voice difficulties restrict my social life" with a "1" (never or rarely). One woman may respond this way because she has a healthy, active social life while another may not experience problems with her voice in social situations because she is socially isolated. Some transsexual women, fearing ridicule or violence, rarely go out in public. As they are not putting themselves into situations where social interaction as a female is difficult the voice may not be perceived as a problem. If such a woman later rates this item with a "2" (sometimes) or "3" (often) it may actually represent progress in her transition.

Ten percent of the spontaneously raised interview content fell under the theme of "Effort and concentration required to produce voice", however only 5% of the content came under the heading, "Physical aspects of voice production". This suggests that for these participants the effort was one of vigilance and constant monitoring of the voice rather than as a sense of physical strain. One participant described her 'feedback loop' which she used to continuously monitor her voice to keep it sounding as she wished. It appears that significant mental effort is commonly needed to achieve and maintain a feminine voice. Advances in the neurosciences may lead to a better specification of that mental effort and new possibilities for intervention.

Alfred Wolfson, the iconic voice teacher from the early 20th century, is quoted as saying, "The voice is the muscle

of the soul.” (Wise, 2007). There is discomfort if the voice does not reflect an individual’s sense of self. The TVQ^{MTF} was developed to assist researchers, speech-language pathologists and transsexual people in an exploration of the voice: how it relates to a sense of gender identity and how it affects day to day function in the world. The current investigation presents initial evidence that the TVQ^{MTF} used thoughtfully and with appropriate knowledge, can present a valid picture of these issues for transsexual women.

References

Brown, G., & Yule, G. (1983). *Teaching the spoken language*. Cambridge, UK: Cambridge University Press.

Carew, L., Dacakis, G., & Oates, J. (2007). The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals. *Journal of Voice*, 21(5), 591-603.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin Company.

Dacakis, G., Davies, S., Oates, J., Douglas, J., & Johnston, J. (2013). The development and initial evaluation of the transsexual voice questionnaire for male to female transsexual persons. *Journal of Voice*, 27(3), 312-320.

Damrose, E. J. (2009). Quantifying the impact of androgen therapy on the female larynx. *Auris Nasus Larynx*, 36(1), 110-112. doi:10.1016/j.anl.2008.03.002.

Davies, S., & Goldberg, J. M. (2006a). *Trans care gender transition: Changing speech*. Vancouver Coastal Health, Transcend Transgender Support & Education Society and Canadian Rainbow Health Coalition.

Davies, S., & Goldberg, J. M. (2006b). Clinical aspects of transgender speech feminization and masculinization. *International Journal of Transgenderism*, 9, 167-196.

Davies, S., Papp, V., & Antoni, C. (in press). WPATH voice and communication standing committee: Companion document.

Gelfer, M., & Schofield, K. (2000). Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male. *Journal of Voice*, 14(1), 22-33.

Günzburger, D. (1993). An acoustic analysis and some perceptual data concerning voice change in male-female transsexuals. *European Journal of Disorders of Communication*, 28, 13-21.

Hancock, A. B., & Garabedian, L. M. (2012). Transgender voice and communication treatment: A retrospective chart review of 25 Cases. *International Journal of Language Communication Disorders*, 48(1), 54-65.

Hancock, A. B., Colton, L., & Douglas, F. (2014). Intonation and gender perception: Applications for transgender speakers. *Journal of Voice*, 28(2), 203-209.

Haynes S., Richard, D., & Kubany, E. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7, 238-247.

Heilmann, J., Miller, J., Iglesias, A., Fabiano-Smith, L., Nockerts, A., & Andriacchi, K. (2008). Narrative transcription accuracy and reliability in two languages. *Topics in Language Disorders*, 28, 178-188.

Hillenbrand, J., & Clark, M. (2009). The role of f0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, and Psychophysics*, 71(5), 1150-1166. doi:10.3758/APP.71.5.1150.

Holmberg, E. B., Oates, J., Dacakis, G., & Grant, C. (2010). Phonetograms, aerodynamic measurements, self-evaluations, and auditory perceptual ratings of male-to-female voice. *Journal of Voice*, 24(5), 511-522.

Jacobson, B. H., Johnson, A., Grywalski, G., Jacobson, A. S. & Benninger, M. S. (1997). The Voice Handicap Index (VHI): Development and validation. *American Journal of Speech Language Pathology*, 6, 66-70.

Lather, P. (1993). Fertile obsession: Validity after poststructuralism. *The Sociological Quarterly*, 34, 673-693.

Messick, S. (1991). Validity of test interpretation and use. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (6th ed.), New York: Macmillan.

McNeill, E. J. M., Wilson, J. A., Clark, S., & Deakin, J. (2008). Perception of voice in the transgender client. *Journal of Voice*, 22(6), 727-733.

Newton, P. (2013). Does it matter what “validity” means? Lecture given at University of Oxford, Department of Education.

Oates, J. (2012). Evidence-based practice in voice therapy for transgender/transsexual clients. In R. Adler, S. Hirsch & M. Mordaunt (Eds.), *Voice and communication therapy for the transgender/transsexual client: A comprehensive clinical guide* (Second ed., pp. 45-68). San Diego, CA: Plural Publishing.

Owen, K., & Hancock, A.B. (2010). The role of self- and listener perceptions of femininity in voice therapy. *International Journal of Transgenderism*, 21(4), 272-284.

Pasricha, N., Dacakis, G., & Oates, J. (2008). Communicative satisfaction of male-to-female transsexuals. *Logopedics Phoniatrics Vocology*, 33, 25-34.

Tannen, D. (1990). *You just don't understand*. NY: Ballantine Books.

Wise, L. (2007). Voice and soul—The Alfred Wolfsohn/Roy Hart Legacy. *Voice and Speech Review*. Voice and Speech Trainers Association. Cincinnati, OH.

World Professional Association for Transgender Health. (2013). Standards of Care: For the Health of Transsexual, Transgender, and Gender-Nonconforming People. Retrieved from http://www.wpath.org/uploaded_files/140/files/Standards%20of%20Care,%20V7%20Full%20Book.pdf

Acknowledgements

Support for this project was provided by a Clinical Research Grant to the first author from the Canadian Association of Speech Language Pathologists and Audiologists. We wish to acknowledge and thank interviewer, Jennifer Wolowic, coder, Jo Nussbaum, and the student transcribers for contributing their expertise and effort. We thank all the transsexual persons who volunteered for our project, especially the five women whose ideas are the substance of this paper.

Authors' Note

A bi-national team consisting of the authors and three researchers from La Trobe University in Melbourne, Georgia Dacakis, Jacinta Douglas and Jennifer Oates collaborated in developing the TVQ^{MTF}. Subsequent research projects have been conducted by different members of the team as indicated by authorship. Copyright to the TVQ^{MTF} is held by Shelagh Davies and Georgia Dacakis. A copy of the Transsexual Voice Questionnaire (Male to Female) may be obtained from this website: www.shelaghdavies.com or by contacting Georgia Dacakis at G.Dacakis@latrobe.edu.au. The questionnaire has no commercial value and is available free of charge.

Correspondence concerning this article should be addressed to Shelagh M. Davies, 2599 West 35 Ave., Vancouver, BC, V6N 2L9 CANADA. Email: sd@shelaghdavies.com.

Appendix

Sample of interview transcription with content coded by theme

J = interviewer; P = MtF transsexual woman

J. Will you just tell me about your voice. What do you think about it?

P. I think it's a, what do I think about my voice? It's probably not where I want it to be./7/

J. Mmm-hmm

P. It's probably, ah, well, for one, I guess everyone's going to recognize you if they listen to you. And I'm just thinking it's probably more of a male voice than a female voice,/3/ and that's what I'm looking at right now, is, having a higher pitch/6/ and I, so that I don't-'caus that's the worst thing you can do, is you can go and look nice, and then you have this deep voice, or just something uh, when you're not thinking about it. And that's the biggest thing, you're always thinking about your voice,/4/ 'caus you don't want to, it not out you basically /3/.