



## Research Article

# Voice Therapy According to the Rehabilitation Treatment Specification System: Expert Consensus Ingredients and Targets

Jarrad H. Van Stan,<sup>a,b,c</sup>  John Whyte,<sup>d</sup> Joseph R. Duffy,<sup>e</sup> Julie Barkmeier-Kraemer,<sup>f</sup> Patricia Doyle,<sup>g</sup> Shirley Gherson,<sup>h</sup> Lisa Kelchner,<sup>i</sup> Jason Muise,<sup>a,c</sup> Brian Petty,<sup>j</sup> Nelson Roy,<sup>f</sup>  Joseph Stemple,<sup>k</sup> Susan Thibeault,<sup>l</sup> and Carol Jorgensen Tolejano<sup>l</sup>

**Purpose:** Clinical trials have demonstrated that standardized voice treatment programs are effective for some patients, but identifying the unique individual treatment ingredients specifically responsible for observed improvements remains elusive. To address this problem, the authors used a taxonomy of voice therapy, the Rehabilitation Treatment Specification System (RTSS), and a Delphi process to develop the RTSS-Voice (expert consensus categories of measurable and unique voice treatment ingredients and targets).

**Method:** Initial targets and ingredients were derived from a taxonomy of voice therapy. Through six Delphi Rounds, 10 vocal rehabilitation experts rated the measurability and uniqueness of individual treatment targets and ingredients. After each round, revisions (guided by the experts' feedback) were finalized among a primary reader (a voice therapy expert) and two external readers (rehabilitation experts outside

the field of voice). Consensus was established when the label and definition of an ingredient or target reached a supramajority threshold ( $\geq 8$  of 10 expert agreement).

**Results:** Thirty-five target and 19 ingredient categories were agreed to be measurable, unique, and accurate reflections of the rules and terminology of the RTSS. Operational definitions for each category included differences in the way ingredients are delivered and the way individual targets are modified by those ingredients.

**Conclusions:** The consensus labels and operationalized ingredients and targets making up the RTSS-Voice have potential to improve voice therapy research, practice, and education/training. The methods used to develop these lists may be useful for other speech, language, and hearing treatment specifications.

**Supplemental Material:** <https://doi.org/10.23641/asha.15243357>

<sup>a</sup>Harvard Medical School, Boston, MA

<sup>b</sup>Massachusetts General Hospital Center for Laryngeal Surgery and Voice Rehabilitation, Boston

<sup>c</sup>MGH Institute of Health Professions, Boston, MA

<sup>d</sup>Moss Rehabilitation Research Institute, Elkins Park, PA

<sup>e</sup>Mayo Clinic, Rochester, MN

<sup>f</sup>University of Utah, Salt Lake City

<sup>g</sup>University of Connecticut Medical School, Farmington

<sup>h</sup>New York University, New York

<sup>i</sup>University of Cincinnati, OH

<sup>j</sup>Emory University, Atlanta, GA

<sup>k</sup>University of Kentucky, Lexington

<sup>l</sup>University of Wisconsin–Madison

Correspondence to Jarrad H. Van Stan: [jvanstan@mgh.harvard.edu](mailto:jvanstan@mgh.harvard.edu)

Editor-in-Chief: Katherine C. Hustad

Editor: Rita R. Patel

Received March 8, 2021

Revision received April 27, 2021

Accepted May 19, 2021

[https://doi.org/10.1044/2021\\_AJSLP-21-00076](https://doi.org/10.1044/2021_AJSLP-21-00076)

For decades, the inadequacy of rehabilitation treatment descriptions has been widely acknowledged (Dijkers et al., 2002; Emerson et al., 1990; Turkstra et al., 2016; Van Stan et al., 2019). Specifically, rehabilitation treatment descriptions are not explicit enough to facilitate (a) replication of efficacy or effectiveness in research or clinical implementation (i.e., the critical pieces of treatment are unclear), (b) comparative effectiveness research (i.e., one cannot obviously know how/if the critical pieces of treatment differ among protocols), or (c) meta-analyses (i.e., which study treatments are sufficiently similar to be combined based on their critical pieces is uncertain). Of note, the terms *efficacy* and *effectiveness* refer to the performance of an intervention under experimentally controlled conditions (i.e., strong internal validity) versus real-world conditions (i.e., strong external validity), respectively.

**Disclosure:** Jarrad H. Van Stan and John Whyte are two of nine copyright holders of the *Manual for Rehabilitation Treatment Specification*. The other authors have declared that no competing financial or nonfinancial interests existed at the time of publication.

The current state of treatment for the most commonly seen voice disorder—nonphonotraumatic vocal hyperfunction (Hillman et al., 2020), also commonly called functional dysphonia (Morrison et al., 1986; Roy & Leeper, 1993) or primary muscle tension dysphonia (Altman et al., 2005; Van Houtte et al., 2011)—can concretely illustrate the problem. There are at least 10 research protocols for such patients that have demonstrated statistically significant improvements in Voice Handicap Index (VHI) ratings (Jacobson et al., 1997) and/or voice quality, including circumlaryngeal massage (Roy et al., 1997; Roy & Leeper, 1993), confidential voice therapy (Verdolini-Marston et al., 1995), conversation training therapy (Gillespie et al., 2019), laryngeal manual therapy (Mathieson et al., 2009), laryngeal reposturing (Roy et al., 2017), two different resonant voice therapy protocols (Roy et al., 2003; Verdolini-Marston et al., 1995), stretch and flow therapy (Watts et al., 2019), vocal function exercises (VFE; Stemple, 2005), and voice production therapy (Behrman et al., 2008). Interestingly, whenever these treatments have been provided in comparative effectiveness studies, they have been found to be noninferior; that is, similar outcomes were found between the compared treatments (Kapsner-Smith et al., 2015; Roy et al., 2001, 2003; Watts et al., 2019). Since it is unknown which commonalities or differences among these treatments are primarily responsible for the positive outcomes, comparative effectiveness studies risk simply comparing treatments composed mostly (or entirely) of the same critical therapeutic pieces (Turkstra et al., 2016; Van Stan et al., 2015, 2019; Whyte & Hart, 2003). For example, all 10 of these protocols are largely made up of the clinician (a) asking the patient to voice repeatedly, (b) providing feedback about vocal performance, and (c) providing cues/information to improve the likelihood that the patient will adhere to the treatment outside of the clinic. To further complicate matters, voice therapy in everyday practice (i.e., standard care) also includes the clinician actions just mentioned. Therefore, when research treatment descriptions are inadequate, frontline clinicians cannot obviously know whether their typical clinical practice aligns with the evidence-based methodology or requires the adoption/adaptation of something new.

There appear to be three major obstacles to improved descriptions of behaviorally based interventions, including voice therapy. First, the critical pieces of treatment are rarely self-evident. Second, interventions are often composed of multiple clinician actions and modified patient functions with minimal indication of how they are connected. Third, the critical pieces of treatment often lack standardized labels and definitions that enable multiple researchers and clinicians to consistently interpret and apply them. The remainder of this article will focus on an expert-consensus approach that addressed these obstacles. Specifically, the consensus approach used the Rehabilitation Treatment Specification System (RTSS; Hart et al., 2019) as the standard to identify and define most (if not all) of the clinician actions and modified patient functions hypothetically responsible for voice treatment outcomes. *The Manual for Rehabilitation Treatment Specification* is available at <https://acrm.org/acrm-communities/rehabilitation-treatment-specification/>

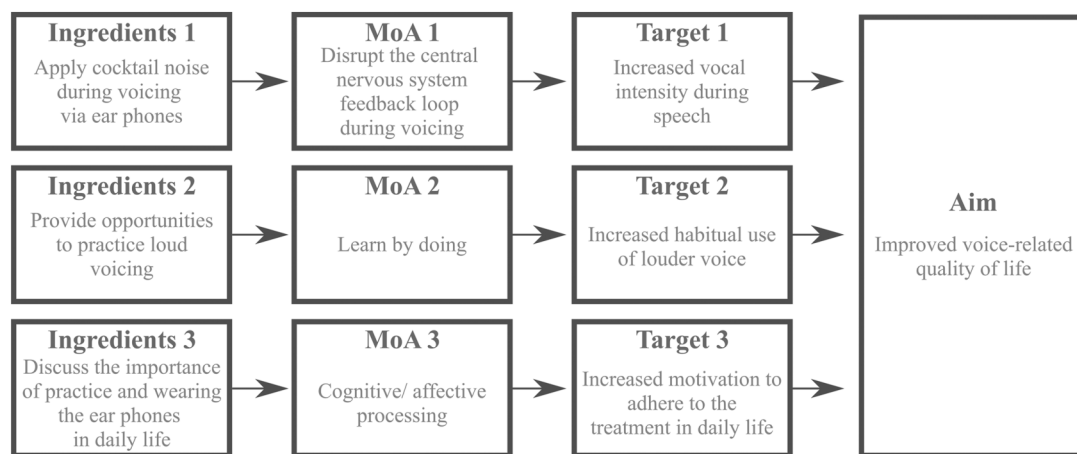
manual-for-rehabilitation-treatment-specification/ and describes the RTSS specification process in detail. In this article, all terms that have definitions in the open-access *Manual* will be underlined when first used.

### ***Identifying the Ingredients and Targets of Treatment***

Vocal rehabilitation treatments are difficult to study because their critical therapeutic pieces are not inherently obvious. For example, clinician actions are often hidden under nonspecific wording that describes what the patient is asked to do (e.g., vocalize), and what the patient is asked to do often contains multiple co-occurring constructs (e.g., voicing necessarily entails breathing and the processing of sensory feedback) with countless variations (e.g., different vowels, consonants, words or short phrases, patterns of musical notes, range of pitches, speed of pitch changes). Furthermore, clinician actions have direct and indirect effects on multiple overlapping patient functions. For example, clinicians cannot alter/influence anterior neck muscle activation levels without likely effecting some change in the patient's central nervous system activation patterns, voice quality, and/or vocal effort. Therefore, vocal rehabilitation treatments require a treatment theory to identify which parts of the intervention are the critical clinician actions, as well as to identify which changes in patient functioning indicate if the clinical actions were successful (Whyte et al., 2014).

The RTSS provides a conceptual framework for specifying any rehabilitation intervention based on a clinician's treatment theory. A treatment theory characterizes how clinician actions are predicted to directly affect patient functioning based on the smallest unit of treatment, the *treatment component*. Figure 1 illustrates the *tripartite structure* of treatment components, which include (a) a singular treatment *target*, the patient function that is to be directly changed by the ingredient(s); (b) one or more *ingredients*, what the clinician does to modify the target; and (c) *mechanism(s) of action* (MoAs), how the ingredient(s) affect(s) the target. According to the RTSS, ingredients and targets must always be observable (i.e., measurable in principle). In contrast, MoAs can be measured or hypothesized because they are typically only measured in research studies to better understand *how* an ingredient affects a target. For example, when a patient is provided opportunities to practice voicing (ingredient) and increases in forward resonance are perceived (target), the clinician hypothesizes that multiple other patient functions have changed in an expected manner (MoAs). Examples of patient functions in the MoA during a semi-occluded vocal tract (SOVT) exercise (Story et al., 2000; Titze, 2006) include impedance matching, the proportion of cricothyroid versus cricoarytenoid muscle activation during voicing (Kochis-Jennings et al., 2014; Lowell & Story, 2006), vocal fold kinematics, or contact patterns during phonation (Patel et al., 2014; Verdolini-Marston et al., 1995). Modified patient functions in the MoAs were not observed/measured, only hypothesized to have occurred because the observed target changed after an observed ingredient was delivered.

**Figure 1.** The tripartite structure of a treatment component and the relationship between treatment components and aims. The arrows point in the direction of causality; that is, the clinician delivers an ingredient (or multiple ingredients) to directly affect a singular treatment target. The mechanism(s) of action is how the ingredient is hypothesized to affect the target. Ingredients do not directly affect aims; aims are indirectly achieved through one or more targets. The hypothetical treatment outlined contains three treatment components to achieve one aim. Treatment Component 1 is within the *Organ Functions* group (top row), Treatment Component 2 is within the *Skills and Habits* group (middle row), and Treatment Component 3 is within the *Representations* group (bottom row). The hypothetical patient has a history of Parkinson's disease, presenting with reduced intelligibility, vocal intensity, and intonation. The patient was prescribed a device that provides cocktail noise via ear buds whenever the patient voices in daily life (Treatment Component 1). During therapy, the patient practiced repetitive loud voicing to increase his vocal intensity without the ear buds (Treatment Component 2). Also, the patient said that he keeps forgetting to practice his exercises and use the ear buds. Therefore, during therapy, the clinician and patient discuss the importance of these two ingredients in hopes of improving adherence (Treatment Component 3). MoA = mechanism of action.



### Connecting Multiple Ingredients to Their Different Targets

Voice treatments can be difficult to meticulously investigate because they are often composed of multiple ingredients (e.g., request the patient to voice repeatedly, provide information regarding vocal hygiene) intended to directly modify different targets (e.g., improved voice quality, perform vocal hygiene strategies as directed in daily life). For example, providing opportunities to practice voicing should more optimally improve voice quality than providing information about the importance of vocal hygiene. Likewise, providing information about the importance of vocal hygiene strategies should more optimally improve patient adherence to (or implementation of) vocal hygiene strategies in daily life than providing opportunities to practice voicing. Therefore, in order to understand the effects of treatment, voice therapy requires theoretically aligning/connecting the multiple ingredients with their specific targets.

To guide clinicians' and researchers' grouping of ingredients with their respective targets, the RTSS divides treatment components into three orthogonal groups: *Organ Functions*, *Skills and Habits*, and *Representations*. *Organ Functions* treatment components change the efficiency of (or replace) organs or organ systems through ingredients that are physical (e.g., apply pressure, surgery), related to challenging organ systems (e.g., resistance training), devices (e.g., tracheoesophageal voice prosthesis, voice amplification), and so forth. *Skills and Habits* treatment components improve mental or behavioral abilities and contribute to the formation of mental/behavioral habits through practice-based

ingredients (e.g., opportunities to practice voicing in a specific way). Of note, *Skills and Habits* treatment components have two subdivisions inspired by the International Classification of Functioning, Disability and Health (World Health Organization, 2001): (a) *function-like skills* (body functions, such as dynamic balance or sustained attention) and (b) *activity-like skills* (specific learned activities, such as activities of daily living), because both of these ostensibly respond to opportunities to practice. *Representations* treatment components change mental representations such as knowledge, attitudes, emotions, beliefs, motivation, and so forth through informational ingredients (e.g., provide vocal hygiene information). Table 1 provides examples for each treatment component group. Finally, as shown in Figure 1, the RTSS defines the aims of treatment as the indirect changes in patient functioning due to changes in one or multiple targets. The critical difference between targets and aims is that targets are hypothesized to be a direct effect of ingredients whereas aims result from achieving one or more targets. Thus, the effects of ingredients on aims are indirect, achieved only through their influence on intermediate targets.

Most vocal rehabilitation treatments require the patient to *volitionally* perform a behavior. Therefore, when creating treatment component groupings (i.e., connecting ingredients with their respective targets), the RTSS asks clinicians to pay careful attention to which treatment components require the patient's active participation/effort (i.e., *volition*; Whyte et al., 2019). To be specific, the RTSS identifies two possible reasons to select and describe an ingredient: (a) The ingredient is hypothesized to directly affect

**Table 1.** Treatment group examples using the Rehabilitation Treatment Specification System (RTSS).

Group	Treatment example	Selected target	Ingredient(s) for selected target
Organ Functions	SpeechVive <sup>a</sup>	Reflexively increased vocal intensity as measured by acoustic sound pressure level (SPL) in decibels (dB)	<ul style="list-style-type: none"> <li>During voicing in daily life, apply multitalker babble noise via a single earphone into the patient's better ear (determined by hearing threshold evaluation). <u>Dose:</u> At an intensity (dB) that results in the patient speaking 3 dB above his/her own comfortable SPL. <u>Progression:</u> None noted in the protocol.</li> </ul>
	Laryngeal manual therapy <sup>b</sup>	Decreased anterior neck muscle activation at rest (no voicing). Measured according to clinician subjective (tactile) judgment.	<ul style="list-style-type: none"> <li>During vegetative breathing (no voicing), apply light circular pressure to the bilateral anterior neck using the pads of the index, second, and third fingers from both hands. <u>Dose:</u> Apply until no further decrease in muscle activation is noted in the targeted area. <u>Progression:</u> Start on the sternocleidomastoids, progressing from the areas of least anterior neck muscle activation to the areas of highest anterior neck muscle activation.</li> </ul>
	Laryngeal reposturing <sup>c</sup>	Reflexively decreased overall dysphonia. Measured according to clinician subjective (auditory perceptual) judgment.	<ul style="list-style-type: none"> <li>During voicing, apply constant inward and downward pressures with the side of the index finger (with hand configured in a "reverse karate chop") starting anteriorly above the hyoid bone, then moving onto the body of the hyoid bone, then the inferior aspect of the hyoid bone, and ending in the thyrohyoid space. <u>Dose:</u> Apply a degree of pressure that is commensurate with the degree of muscle activation/resistance encountered in the underlying muscles, and to overcome or resist the action of those muscles. Apply the pressure to an area until voice quality improvement plateaus or improves to normal. <u>Progression:</u> None noted in the protocol.</li> </ul>
	Expiratory muscle strength training <sup>d</sup>	Increased maximal expiratory pressure (cm H <sub>2</sub> O)	<ul style="list-style-type: none"> <li>Patients forcefully exhale into an expiratory muscle strength trainer. <u>Dose:</u> Device set at a resistance level equal to 75% of the participants' maximum expiratory pressure before treatment. Patients perform 5 repetitions, 5 times per day, for 4 weeks. <u>Progression:</u> None noted in the protocol.</li> </ul>
Skills & Habits	Vocal function exercises <sup>e</sup>	Sustain phonation for a specific duration of time (seconds) based on vital capacity (liters) divided by a desired mean airflow (0.08 L per second).	<ul style="list-style-type: none"> <li>Provide opportunities to practice a template of voicing (i.e., sustained voicing until completely run out of air, on an /o/ buzz, at soft intensity, at specific musical notes, and using abdominal breathing). <u>Dose:</u> Patient performs 10 trials, 2 times per day for 4–6 weeks. <u>Progression:</u> None noted in protocol.</li> </ul>
	Circumlaryngeal massage <sup>f</sup>	Increased accuracy switching between disordered and normal voicing ("negative practice"). Measured according to clinician subjective (auditory perceptual) judgment.	<ul style="list-style-type: none"> <li>Provide feedback on phonatory duration (in seconds) after a trial. <u>Dose:</u> 100% frequency (after every trial). <u>Progression:</u> None noted in protocol.</li> <li>Provide volition ingredients: Cues to correctly perform the exercise template (i.e., /o/ buzz, soft intensity, maintain requested musical note, use abdominal breathing, keep sustaining the trial until the patient runs out of air). <u>Dose:</u> As needed, whenever a part of the template is not performed correctly. <u>Progression:</u> None noted in the protocol.</li> <li>Provide opportunities for the patient to practice switching between their own disordered voice and their new normal voice during automatic speech (e.g., days of the week) and/or the Rainbow Passage. <u>Dose:</u> Repeat until the patient is ~100% accurate at quickly switching between voices. <u>Progression:</u> Provide cues to switch between voices more quickly/frequently as patient improves.</li> </ul>

(table continues)

**Table 1.** (Continued).

Group	Treatment example	Selected target	Ingredient(s) for selected target
Representations	Vocal hygiene <sup>g,h</sup>	Increased knowledge of vocally healthy and unhealthy behaviors. Measured via patient recall of the information.	<ul style="list-style-type: none"> <li>• Provide feedback immediately after only inaccurate productions (e.g., “that’s not it” or “fix it”). <u>Dose:</u> 100% frequency (after every inaccurate switch). Fading occurs naturally as inaccurate productions no longer occur. <u>Progression:</u> None noted in the protocol.</li> <li>• Provide volition ingredients: Cues to correctly perform the negative practice (i.e., maintain attention to how the voice sounds/feels, provide a vocal model of patient’s disordered or normal voice). <u>Dose:</u> As needed, whenever an aspect of the negative practice is not performed correctly. <u>Progression:</u> None noted in the protocol.</li> <li>• Oral and written information was provided to the patient regarding vocally healthy and unhealthy behaviors, as well as strategies for modifying or eliminating unhealthy vocal behaviors. <u>Dose:</u> Information was repeated until patient accurately recalled it at 100% accuracy. <u>Progression:</u> None noted in the protocol.</li> <li>• Provide MP4 players that included videos of the clinician performing exercises, patient “self-as-model” examples, and a peer testimonial of patients with good therapy outcomes. <u>Dose:</u> To be reviewed at every practice session outside of the clinic. <u>Progression:</u> None noted in the protocol.</li> <li>• Provide cues to the patient to talk about the positive changes he/she has made in the past. <u>Dose:</u> As needed, when the patient produces statements indicating low self-efficacy. <u>Progression:</u> None noted in the protocol.</li> </ul>
	Video-enhanced voice therapy <sup>i</sup>	Increased likelihood that the patient will practice prescribed exercises 5 times daily. Measured via patient self-report.	
	Motivational interviewing in voice therapy <sup>j</sup>	Modified beliefs regarding the patient’s ability to change their vocal behavior (i.e., increased voice-related self-efficacy)	

<sup>a</sup>Stathopoulos et al. (2014). <sup>b</sup>Mathieson et al. (2009). <sup>c</sup>Roy et al. (2017). <sup>d</sup>Pitts et al. (2009). <sup>e</sup>Stemple (2005). <sup>f</sup>Roy et al. (1997). <sup>g</sup>Roy et al. (2001). <sup>h</sup>Roy et al. (2002). <sup>i</sup>van Leer & Connor (2012). <sup>j</sup>Behrman (2006).



patient functioning (i.e., it affects a *direct target*), or (b) the ingredient is hypothesized to increase the likelihood of the patient performing a therapy activity as directed (i.e., it affects a *volition target*). For example, instructional and motivational ingredients (*volition ingredients*) are ostensibly required for patient adherence to the prescribed treatment, such as *correctly* performing a vocal exercise *twice a day* (a *volition target*). When the patient is adherent, the prescribed treatment—such as the vocal exercise (ingredients for the direct target)—can improve voicing in some way (direct target). To help clinicians consistently think about and specify the volition ingredients they provide, the RTSS uses a theoretical framework from health psychology that considers the patient’s Capability, Opportunity, and Motivation necessary for Behavior change (Michie et al., 2011). Essentially, Capability (the individual’s psychological and physical capacity to engage in a behavior), Opportunities (the factors that lie outside the individual that make the behavior possible or prompt it), and Motivation (the brain processes that energize and direct behavior) are overlapping/interacting components that generate a desired behavior. Although Capability, Opportunity, and Motivation can be considered separately, they are not intended to be conceptually unique categories because all individual volition ingredients can affect behavior in multiple ways.

### ***Creating Standard Ingredient and Target Labels***

The third obstacle to improved treatment research is that any individual treatment ingredient or target identified through any process needs to be defined and operationalized (i.e., standardized). Without standardization, the same treatment theory could be described in different words, or different treatment theories could be described in the same, or very similar, words. To the authors’ knowledge, the most comprehensive approach attempting to describe vocal rehabilitation treatments in a standardized manner is the Taxonomy of Voice Therapy (Van Stan et al., 2015). Its authors reviewed clinical documentation, research treatment protocols, and voice therapy textbooks to identify a comprehensive and standardized set of labeled and defined therapy tasks (termed *tools* in the taxonomy). Then, existing treatment theories from the field of motor control and learning were used to develop a framework that grouped therapy tasks along quasi-orthogonal dimensions. Despite standardized terminology and a theoretically motivated framework, the resulting items were a mix of ingredients (e.g., digital manipulation), targets (e.g., loudness modification), or a combination of both, for example, held positions (an ingredient) for lengthening muscle (a target). Also, since the underlying theory and framework were based solely on a view toward voice-related behavior changes, the interpretability of the taxonomy for use in a broader context (e.g., other domains of speech-language pathology or rehabilitation) was limited.

In order to standardize the naming and operationalizing of individual ingredients and targets, all of the different

labels that are used to describe treatment need to be identified and then classified as unique (i.e., an individual ingredient or target) or redundant (i.e., grouped together to form an individual ingredient or target). Theory-driven uniqueness judgments are necessary to arrive at the smallest number of unique concepts (either target or ingredient) that cover a treatment domain, such as voice therapy. Once the putatively unique concepts are identified, all the different labels that address each concept can be systematically organized under one label and operationalized. However, until recently, the RTSS had not explicitly operationalized the concepts of unique and redundant. Therefore, we developed a methodology that formally describes when concepts can be considered unique or redundant (Van Stan et al., 2020). The primary purpose of this study was to use that methodology to produce a comprehensive list of voice treatment ingredients and targets (content validity) that adhere to the RTSS framework (construct validity) and has supramajority ( $\geq 80\%$  agreement) expert consensus (face validity). Of note, the goal of the study was not to provide a catalog of only targets and ingredients for which there is evidence-based support. It is hoped that a preliminary list of measurable and unique targets and ingredients will provide a useful starting point for future RTSS-descriptions of voice interventions. Once treatments are described in a standard way, clinicians and researchers can identify the key similarities and differences among them and begin to test their hypotheses regarding which clinician actions are more or less efficacious/effective for changing various patient functions, which will also help to guide the identification of underlying MoAs.

### **Method**

A detailed discussion of the rationale, limitations, and advantages of the method used here is beyond the scope of this article. Such details can be found elsewhere (Van Stan et al., 2020).

### ***Delphi Process***

The Delphi process (Linstone & Turoff, 1975) was used because it is a systematic way to obtain consensus on a topic from a panel of independent experts. It is also beneficial for investigating issues with minimal empirical data, such as the black box of rehabilitation treatments (Whyte & Hart, 2003). Generally, the Delphi process consists of structured questionnaires across multiple rounds of feedback. Questionnaire content is revised between rounds to improve and/or represent areas of agreement/disagreement. Key characteristics of the method include (a) anonymity and independence of the participating experts to reduce bias, (b) structured questionnaires for each round to collect uniformly structured data, (c) regular feedback among experts to help experts understand and respond to others’ feedback, (d) a facilitator to synthesize the results of each round and plan the next round, and (e) external readers to provide feedback on the facilitator’s data synthesis and minimize facilitator

bias. Regarding anonymity, all ratings and feedback from the voice therapy experts (except for the in-person Round 2, as explained in Data Acquisition section below) were de-identified by the facilitator. Then, the expert responses were grouped (not maintained as individual raters) before the two readers, and subsequently the voice therapy experts reviewed each round's results and recommendations. The Delphi structure and content used here were based on a previously successful effort to produce a treatment taxonomy in the field of psychology (Michie et al., 2013; Wood et al., 2015).

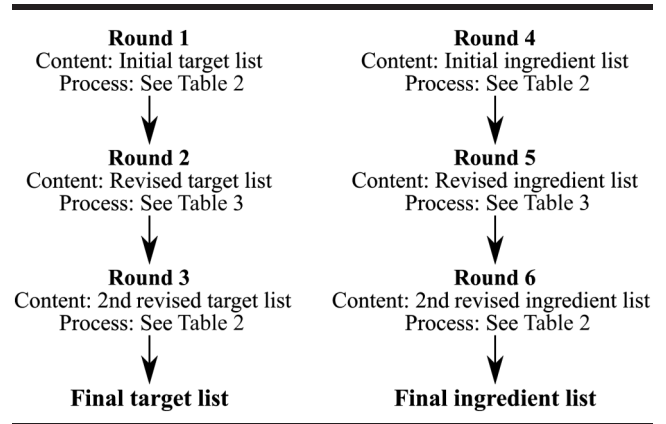
### Participants

A total of 13 experts participated in the Delphi Rounds. Ten expert voice-specialized speech-language pathologists (SLPs) across eight voice centers participated in six Delphi Rounds (Authors 3 through 13). These experts were recruited to cover a range of expertise across various types of voice disorders (e.g., neurological, behavioral, structural), therapeutic approaches (e.g., manual voice therapy, resonant voice therapy, respiratory training), and type of clinical practice (five experts were clinical researchers, and five experts were frontline clinicians who provide voice therapy regularly). Expertise in specific voice disorders and therapeutic approaches for each participant was determined via peer-reviewed articles for the five clinical researchers and involvement in published studies, educational trainings, and clinical caseload for the five frontline clinicians. Ten voice experts were chosen, as opposed to a higher or lower number, for two reasons: (a) It was desirable to equally represent clinical researchers and frontline clinicians (an even number was necessary), and (b) saturation in the diversity of clinical expertise (see criteria in the previous sentence) was met with 10 total participants. There was one facilitator throughout the Delphi process who is an SLP specialized in voice disorders and part of the teams that developed the taxonomy of voice therapy and the RTSS (J. V. S.). To reduce the risk of facilitator bias, two external readers participated who were not voice therapy experts; one is a physiatrist specialized in traumatic brain injury rehabilitation who led the development of the Rehabilitation Treatment Taxonomy (Hart et al., 2014) and the RTSS (J. W.), and the other is an SLP specialized in neurologic speech and language disorders (J. R. D.).

### Data Acquisition

Figure 2 provides an overview of the data collection and analysis steps in flowchart format. The initial lists of voice therapy ingredients and targets were derived from the taxonomy of voice therapy, because the taxonomy was developed from a comprehensive literature search and investigation of everyday clinical notes. The ingredients (termed *tools*) in the taxonomy were reframed into lists of ingredients and targets according to the RTSS concept of a treatment component group (i.e., *Organ Functions*, *Skills and Habits*, *Representations*). The list of targets was mainly constrained to two groups of treatment components that

**Figure 2.** Flow chart depicting each Delphi Round, the contents that were evaluated each round, and the process used to evaluate each round's content. The arrows represent revisions to the list that were completed by the two readers and one facilitator according to expert feedback.



have a theoretically finite number of targets: (a) functions of the respiratory system, larynx, resonators, and sensorimotor system (*Organ Functions*); and (b) performing those functions in a skilled or habitual way that usually involves hierarchical learning/ generalization (*Function-like Skills and Habits*). Two other treatment groups have a theoretically infinite number of targets and, therefore, were mostly excluded from this project: *Activity-like Skills and Habits* and *Representations*. That is, there are unlimited numbers of potential skilled activities (e.g., voicing during presentations, while socializing at the bar, acting, singing a specific song) and mental states (e.g., changes in beliefs, affect, likelihood of doing something). Then, the ingredients list was limited to include only the clinician actions that are commonly used to affect the identified *Organ Functions* and *Function-like Skills and Habits* targets. Some ingredients and targets from the *Representations* treatment components were included in the lists to represent the prevalent and presumably important vocal hygiene part of voice therapy (Behlau & Oliveira, 2009; Behrman et al., 2008; Holmberg et al., 2001; Roy et al., 2001, 2002). Since *Activity-like Skills and Habits* treatment components were excluded, the resulting ingredients and targets list represented general behaviors (e.g., opportunities to practice voicing to increase forward resonance), rather than specific activities (e.g., opportunities to practice increased forward resonance during a work presentation). Since *Representations* treatment components were generally excluded (with the exception just noted regarding vocal hygiene), the lists were not expected to comprehensively capture broad methods to convey information (e.g., teaching methods), the specific content (e.g., information bits) conveyed with those methods, nor the mental states to be changed by those ingredients (e.g., changes in beliefs, affect, likelihood of doing something). The facilitator and external readers iteratively revised the initial lists of ingredients and targets until all three investigators agreed that the lists accurately aligned with the RTSS framework.

The Delphi process consisted of six rounds. The experts evaluated the target list first (Rounds 1–3) and the ingredients list second (Rounds 4–6). All rounds were completed online via REDCap surveys, except for Round 5, which was completed during an in-person meeting. In Rounds 1 (targets) and 4 (the first ingredients round), the experts answered the standard questions in Table 2 for each target or ingredient. In Rounds 2 (targets) and 5 (ingredients), the experts answered probe questions in Table 3 regarding targets or ingredients that did not reach consensus in the previous round (1 or 4, respectively). These probe questions addressed ambiguities in the experts' judgments of uniqueness or redundancy in Round 1 or 4. In Rounds 3 and 6, the experts again answered the standard questions in Table 2 for the targets or ingredients that had not yet reached consensus. After Round 6, the facilitator created a list of ingredients and targets that were provided to the experts and readers for final approval.

The group approached the Round 5 ingredient probe questions in Table 3 during an in-person meeting. The group met in person because many of the experts found it difficult to comprehensively express their feedback in writing (on average, experts spent over 8 hr of work per round), and it was felt that multiple agreement issues could be

resolved faster during an interactive conversation. The meeting started with didactic orientation to the RTSS and then focused on four topics covering the groups of ingredient labels with the most overlap. The topics were discussed in small groups (three or four experts per group) for 30 min, followed by a full group discussion for 60 min during which the entire group explicitly agreed on the areas of consensus. The small groups were composed of one reader or facilitator and three or four randomly assigned vocal rehabilitation experts. Each discussion topic was guided by a handout containing a description of the overlap among ingredient labels and probe questions filled in with specific ingredients, targets, diagnoses, impairments, and severity level of impairments.

### Data Analysis

The observability and conceptual uniqueness ratings for each target or ingredient were examined after every round. When eight out of 10 experts agreed that a target or ingredient was both measurable and conceptually unique—or agreed that multiple targets or ingredients were redundant and/or not measurable—consensus was considered to have been achieved. After each round, expert comments were used to guide revisions of labels and underlying definitions

**Table 2.** Delphi Round 1 and 3 questions for each target; Round 4 and 6 questions for each ingredient.

Questions presented for each target or ingredient		Response options
0	Do you use this [target or ingredient] in your practice and/or research?	Binary: yes, no
1	Can this [target or ingredient] be observed and measured empirically (whether objectively or subjectively)?	Categories: definitely no, probably no, not sure, probably yes, definitely yes Paragraph free text box
2	What subjective or objective measure would you use to quantify this [target or ingredient] during clinical care? (Ask if Question 1 answer: "probably yes" or "definitely yes")	Paragraph free text box
3	Why are you unsure, or do not think, that this [target or ingredient] is observable/measurable? (Ask if Question 1 answer: "definitely no," "probably no," or "not sure")	Paragraph free text box
4	Is this [target or ingredient] conceptually unique or redundant compared to other [targets or ingredients]?	Binary: unique, redundant
5	Redundant with what other [target or ingredient] and why? (Ask if Question 4 answer: "redundant")	Paragraph free text box
6	How would you revise these [targets or ingredients] to reduce or eliminate redundancy? (Ask if Question 4 answer: "redundant")	Paragraph free text box
7	If your opinion is that redundancy cannot be reduced or eliminated, which [target or ingredient] is preferable and why? (Ask if Question 4 answer: "redundant")	Paragraph free text box
8	Is the definition of this [target or ingredient] incorrect in any way?	Binary: yes, no
9	What is incorrect in this [target's or ingredient's] description and how would you suggest addressing it? (Ask if Question 8 answer: "yes")	Paragraph free text box
10	Is the definition of this [target or ingredient] missing any critical dimensions or information?	Binary: yes, no
11	What is missing in this [target's or ingredient's] description and why is it important to include this additional information? (Ask if Question 10 answer: "yes")	Paragraph free text box
12	Is there any redundancy in the definition of this [target or ingredient]?	Binary: yes, no
13	How would you revise the description to minimize or eliminate the redundancy? (Ask if Question 12 answer: "yes")	Paragraph free text box
14	Are there any [targets or ingredients] not included here? (Ask after rating all targets or ingredients)	Binary: yes, no
15	Please describe any [targets or ingredients] that you did not see included. (Ask if Question 14 answer: "yes")	Paragraph free text box



**Table 3.** Target and ingredient probe questions.

Target probe questions – Delphi Round 2	Specific targets for Patient Functions A and B
<p>Question 1: If your clinical priority is to change Patient Function A, would your treatment plan differ in any way from a situation where your clinical priority is to change Patient Function B?</p> <p>a. <i>If yes</i>, the patient functions are unique. Please state the ingredients and dose of ingredients you would use for both targets, emphasizing how they would be different.</p> <p>b. <i>If no</i>, the patient functions are redundant. Please state which target phrasing is preferable and what ingredients (as well as dosing of the ingredients) you would use for the target.</p>	<p>i. Improve a patient’s accuracy of self-monitoring their habitual loudness/Decrease the patient’s habitual loudness</p> <p>ii. Increase a patient’s habitual loudness/Increase the patient’s habitual pitch</p> <p>iii. Decrease a patient’s overall dysphonia/Decrease the patient’s habitual pitch</p> <p>iv. Decrease a patient’s breathiness/Decrease the patient’s roughness</p> <p>v. Decrease a patient’s strain/Decrease the patient’s roughness</p> <p>vi. Increase flow phonation/Increase breathy glottal onsets</p>
<p>Question 2: Despite the provision of very similar (or the same) treatment ingredients, are there clinical situations when one change in the patient (Patient Function A) would define treatment success and other times when a more proximal or distal change in the patient (Patient Function B) would define treatment success?</p> <p>a. <i>If yes</i>, the patient functions are unique. Please describe a clinical situation in which each underlined target defines clinical success. Please include what ingredients (and dose of those ingredients) you would use for the target(s).</p> <p>b. <i>If no</i>, then one of these patient changes is always a means to the end of achieving the other patient function; i.e., they are redundant. In your opinion, what target defines clinical success in patients with changes in one or both functions? Also, what ingredients (and dose of ingredients) would you use to achieve this target?</p>	<p>i. Decreased muscle activation/Increased passive range of motion</p> <p>ii. Improved alignment/Decreased muscle activation</p> <p>iii. Decreased muscle activation/Decreased pain/soreness/discomfort</p> <p>iv. Improved kinesthetic awareness/Improved resonance</p> <p>v. Improved abdominal movement/improved respiratory coordination</p>
<p>Question 3: Although many treatment targets may improve aspects of patient functioning that contribute to the aim of Patient Function A, can you provide treatments that directly target Patient Function A?</p> <p>a. <i>If yes</i>, this patient function can be a target. Please state the ingredients and dose of ingredients you would use for improving Patient Function A.</p> <p>b. <i>If no</i>, improving Patient Function A is too broad to be directly changed by a treatment ingredient(s); i.e., it may always be an aim. Please provide two targets and their associated ingredients (as well as dose of ingredients) that could be directed at the aim of improved Patient Function A.</p>	<p>i. Improved respiratory support</p> <p>ii. Decreased vocal effort</p> <p>iii. Decreased pain/soreness/discomfort</p>

(table continues)

**Table 3.** (Continued).

Ingredient probe questions – Delphi Round 5	Specific ingredients for Clinician Actions A, B, C, etc.
<p>Question 1: Would the patient functions you hope to change during treatment be different when you perform Clinical Action A versus Clinical Action B?</p> <p>a. <i>If yes</i>, the clinician actions are unique. For each ingredient, please state the patient function(s) (i.e., target) that is directly affected by that ingredient, how you would use the ingredients for their respective target(s), and why the ingredient–target relationships are different.</p> <p>b. <i>If no</i>, the clinician actions are redundant. Please state which ingredient phrasing is preferable, what target(s) you would affect with the ingredient, and how you would modify the ingredient based on patient characteristics (diagnosis, severity, etc.).</p>	<p>i. Pitch glides, loudness glides</p>
<p>Question 2: When you perform Clinical Action A, B, or C, is there an underlying “common denominator” among them that you believe contributes to changes in patient functioning?</p> <p>a. <i>If yes</i>, some or all clinical actions may be redundant.</p> <p>i. Please describe the common denominator across the ingredients as well as how that common denominator may be varied in manner or amount/dose.</p> <p>ii. Although there is a common denominator among these ingredients, what is the significance of the individual variations themselves? In other words, why would a clinician choose to use ingredient X (Version A) versus ingredient X (Version B)?</p> <p>b. <i>If no</i>, then the clinician actions are unique. Please state the targets each ingredient affects and how each ingredient may be varied in manner or amount/dose, emphasizing how these ingredient–target pairings are different.</p>	<p>i. Semi-occluded vocal tract, inverted megaphone, siren, lip trill, tongue trill, lip-tongue trill, exhale through a constriction, gargling</p> <p>ii. Apply manual stretch, traction, touch, pressure, thyroid pull down, hyoid pushback, shaking</p>
<p>Question 3: Although many patient functions could be affected by Clinical Action A, would you use Clinical Action A differently for Patient Function A versus Patient Function B*?</p> <p>*While the “Clinical Action A” was specifically filled in before answering the question, the “Patient Functions A, B, C, etc.” were left to the expert to fill in according to their treatment theory.</p> <p>a. <i>If yes</i>, Clinical Action A is multiple ingredients. What differences are there in the manner or amount/dose of Ingredient A when used for Target A versus Target B?</p> <p>b. <i>If no</i>, Clinical Action A is a unique ingredient. What are the different ways you have varied Ingredient A in manner or amount/dose to improve its effect on Target A and/or Target B?</p>	<p>i. Pitch glides</p> <p>ii. Loudness glides</p> <p>iii. Phonation duration</p>

for the labels. Once the facilitator incorporated all expert feedback, the revised ingredients and/or targets list was sent to the two external readers for feedback. The facilitator then revised the documents according to the external reader comments. The subsequent round of expert feedback could not begin until all materials were considered final by the facilitator and two external readers.

The goal of this project was to identify measurable and unique targets and ingredients that can represent any voice clinician's treatment theory. In other words, the goal was not to identify every individual target or ingredient that is, or could be, part of voice therapy, but to identify all of the unique concepts that are used to describe individual targets or ingredients in voice therapy. This is an important distinction, because any target or ingredient label can be unique depending on the level of specificity. The purpose of the project was to characterize all voice therapy treatment theories, and this dictated which level of specificity was necessary for the final lists of targets and ingredients. For example, the following ingredients could be individually judged for measurability and uniqueness: *practice maximum sustained voicing at maximum loudness and highest possible pitch* (Ramig et al., 1996), *practice sustained voicing equal to vital capacity divided by 80 ml/s at softest loudness and a middle C* (Stemple, 2005), as well as practically any other possible combination of voicing at a desired duration, loudness, and/or pitch. While this level of specificity is necessary for a well-described treatment protocol, representing/judging every possible individual ingredient at this level of detail is intractable and does not explicitly point out the differences in clinicians' treatment theories. In this example, it appears that clinicians' treatment theories provide the patient with *opportunities to practice voicing* and choose different durations of voicing, loudness levels, and/or pitches depending upon the treatment target. Similar to how a singular pharmacological ingredient can be used in different amounts (e.g., more or less milligrams of aspirin) to affect different targets (e.g., pain or blood clotting, respectively). Therefore, the final target and ingredient lists (instead of representing every individual formulary of a target or ingredient) characterize treatment theories through unique target or ingredient *categories*, and these *categories* are operationalized with underlying ways in which individual targets or ingredients can vary, for example, *opportunities to practice voicing* (ingredient category) *at a loudness level, pitch, and/or for a desired duration* (ingredient category delivered in a specific way).

## Results

This project resulted in a new tool called the RTSS-Voice, which consists of 35 target categories and 19 ingredient categories that were thoroughly vetted and agreed upon by 13 experts adhering to the RTSS framework. The final lists that make up the RTSS-Voice can be found in Appendixes A and B, respectively. Supplemental Materials S1 and S2 illustrate the expert ratings of measurability and conceptual uniqueness for all targets and ingredients (respectively) across the Delphi-type exercises, as well as how each

target and ingredient label was modified throughout the process. Of note, many target and ingredient categories are labeled according to perceptual terms (e.g., pitch, loudness) instead of objective terms (i.e., fundamental frequency, vocal intensity). This is because changes in these targets, or the correct delivery of these ingredients, were most often (a) evaluated in real time according to the clinician's perception and (b) objective measures were used only if they were shown to be a strong quantitative correlate of a reliable and valid perception. Within the Results section, a general summary of each Delphi Round (i.e., expert consensus and overarching changes inspired by each round) will be presented. Second, concrete examples of how individual treatment target or ingredient labels evolved throughout the process will be provided, as well as the underlying rationale for the evolution.

## Targets

During Round 1, the concept of uniqueness resulted in less consensus across the 29 target labels than the concept of measurability; median absolute agreement was six experts versus nine experts, respectively. The vocal rehabilitation experts reported that most patient functions overlapped in some way. They also did not agree regarding how much or what type of overlap could be deemed irrelevant to uniqueness judgments or to justify redundant judgments. To address this uncertainty, we formally operationalized the concepts of unique and redundant through the development of three probe questions in Table 3 (Van Stan et al., 2020). Target labels with the most overlap among each other from Round 1 ratings were used to create variations of the probe questions. Revisions to the targets list after Round 1 included (a) wording modifications to remove or improve unclear statements, (b) additional or revised objective and subjective measures to quantify the target labels, and (c) creation of 15 new target labels that were identified by experts as missing, or were split and reorganized from seven existing target labels. Because of the significant target label restructuring (adding, splitting, combining of target labels) and the uncertainty in how overlap related to uniqueness, all target labels were judged again for uniqueness in Round 2 regardless of supra-majority consensus in the previous round.

Target labels achieving supramajority consensus on both observability and uniqueness/redundancy increased after Round 2 compared to Round 1: 68% (25 of 37) versus 27% (six of 29), respectively. Also, agreement on uniqueness increased for the 22 target labels included in both Round 2 and Round 1; median absolute agreement increased from seven to eight experts. Revisions to the list of targets after Round 2 were much more limited than revisions after Round 1. Round 2 revisions included (a) modifications of definitions and descriptions of individual target labels to increase uniqueness and decrease redundancy and (b) adding four new target labels that experts stated were missing and/or resulted from splitting/reorganizing four existing target labels.

All 37 target labels achieved supramajority consensus on both observability and uniqueness or redundancy after

Round 3. Expert feedback from Round 3 resulted in the deletion of two target labels (i.e., overall dysphonia and vocal effort), and revisions to the labels and definitions were completed to maximize conceptual uniqueness, measurability, and general clarity. The final list consisted of 35 target labels that achieved supramajority consensus on uniqueness and measurability. The next few paragraphs will describe how the probe questions guided the development of consensus on specific issues throughout Rounds 2 and 3.

### Target Probe Question Examples

The variations of Probe Question 1 produced consensus rationales relying on the following principle: Multiple physiologically correlated patient functions are unique targets if each individual patient function is optimally improved by different ingredients or the same ingredients delivered differently. The following examples will concretely illustrate this principle:

- Are there differential treatments for *improving motor output* (e.g., *decreased strained voice quality*) and for *improving sensory discrimination* (e.g., *increased discrimination of various levels of strained voice quality*)? Yes, these two can be unique treatment targets despite their sensorimotor codependence; that is, voice production is necessarily refined by auditory feedback during voicing (Bauer et al., 2006; Burnett & Larson, 2002). For example, a patient with bilateral vocal fold nodules needing to *decrease her strained voice quality* would hypothetically be optimally affected by *opportunities to practice voicing in a manner associated with less strained voice quality*, for example, forward resonance and increased mean airflow during voicing (Roy et al., 2003; Verdolini-Marston et al., 1995; Watts et al., 2015). However, patients will often produce changes in strained voicing (and other voice qualities or resonances), but state that they did not hear or feel any difference in their phonation. Therefore, the target of *improved discrimination of various levels of strained voice quality* may also be necessary, which would ostensibly be optimally improved by *opportunities to practice discriminating among various levels of strained voice quality* (e.g., Gartner-Schmidt et al., 2016).
- Are there differential treatments for changes in *habitual loudness*, *habitual pitch*, and *habitual voice quality*? Yes, these can be unique treatment targets despite their physiological correlation and co-occurrence during the act of voicing, for example, *increases in habitual loudness* are often accompanied by some degree of *increases in habitual pitch*. For example, *increased habitual pitch* for a patient transitioning from male to female (with the aim of sounding more feminine) would be optimally attained with *opportunities to practice voicing with increased pitch* (e.g., Hancock & Garabedian, 2013; McNeill, 2006), and the target of *increased habitual loudness* for a patient with presbyphonia would be optimally attained with *opportunities to practice voicing with increased loudness* (e.g., Ziegler et al., 2014).
- Are there differential treatments for changes in multiple auditory-perceptual constructs, for example, *decreased strain*, *breathiness*, *roughness*, and *improved overall dysphonia*? Some auditory-perceptual constructs appear to be unique treatment targets. Specifically, *decreased strain*, *breathiness*, or *roughness* was judged as unique, but *overall improved voice quality* appears to always be an aim (i.e., not a target). This is because the type of voice quality change (in the context of the patient's voice disorder) was closely linked to which ingredients were chosen. For example, if a patient with nonphonotraumatic vocal hyperfunction produces primarily strained (and minimally breathy) voice quality, this could theoretically require *opportunities to practice voicing with increased mean airflow*. However, in contrast, if a patient with the same diagnosis produces primarily breathy (and minimally strained) voicing, this could presumably require *opportunities to practice voicing with decreased mean airflow* (e.g., Gillespie et al., 2013; Gilman et al., 2019). Since the patient could be producing the same level of *overall dysphonia* in both contexts, *overall dysphonia* was not directly affected by ingredients and was likely an aim of treatment. In other words, the experts did not identify an ingredient that was hypothesized to directly affect *overall dysphonia*. Instead, *overall dysphonia* appeared to be indirectly affected by how ingredients change the targets of *breathiness*, *strain*, and/or *roughness*.
- Are there differential treatments for changes in *glottal onsets* and *flow phonation*? Yes, these two can be unique target labels. This is because the two patient functions appear to be modified by different treatment ingredients. For example, to *decrease a pressed glottal onset* in a patient with a unilateral vocal fold polyp, the experts provided ingredients such as *opportunities to practice initiating voicing with /h/*, *opportunities to practice alternating between initiating voicing with /h/ and /ʔ/*, and *opportunities to practice initiating voicing with increased breathiness*. To achieve the target of *flow phonation* in the same patient, the experts concurred with the ingredient of *opportunities to practice voicing with increased mean airflow*.
- Are there differential treatments for changes in *passive or active range of motion* and *muscle activation*? Yes, these two can be unique target labels. *Passive* (or *active*) *range of motion* would be a unique target in the presence of a structural pathology preventing normal movement range. For example, consider a patient whose main issue is postradiation fibrosis of the mandible (i.e., trismus) and no excess muscle activation in the head/neck. In this case, a clinician can target *increased passive* (or *active*) *range of motion of the mandible* without the changes in range of motion being attributable to *reductions in muscle activation*. However, when elevated levels of muscle activation are the primary observation—such as a patient with vocal tremor exhibiting excessive anterior neck muscle activation levels at rest (no voicing)—changes in *passive range of motion*



may be a way of measuring the clinician's target of *decreased muscle activation*. The clinician could *apply pressure to the patient's anterior neck during rest* to presumably *increase lateral passive range of motion of the hyoid bone at rest*. Any observation of *increased passive range of motion of the hyoid bone* would be an estimate of the target *reduction in muscle activation at rest* (e.g., Mathieson et al., 2009).

The variations of Probe Question 2 produced consensus rationales that relied on identifying if causally linked patient functions were always a means to an end (i.e., in the MoAs) or could be an end unto themselves (i.e., a target). According to the RTSS, treatment ingredients are considered successful based on whether the target (i.e., singular, hypothesized patient function) is modified as desired. When a modified patient function is the arbiter of an ingredient's success, it is the ingredient's target, or an end unto itself. When a modified patient function is not the ingredient's marker of success or failure, the function is within the ingredient's MoA, or a means to an end. The following examples will concretely illustrate the subtleties of this principle:

- Since the patient functions of *muscle activation* and *strained voice quality* are sequentially linked, are there contexts where each of these patient functions is the criterion for a treatment's success? Yes, meaning that both target labels are unique. Consider a patient with unilateral vocal fold paralysis (after successful vocal fold medialization) exhibiting elevated anterior neck muscle activation at rest and moderately strained voice quality. Evidence has shown that *providing manual pressure to the anterior neck in the absence of voicing* can *reduce muscle activation at rest* (Mathieson et al., 2009). However, if the patient continues to produce strained voicing after reducing her muscle activation at rest, evidence has shown that *reductions in strained voice quality* can occur when the clinician *applies manual pressure to the anterior neck during voicing* (Roy et al., 2017). In this context, changes in *muscle activation* would be in the MoAs (i.e., a means to an end) when the clinically desired change in patient function (i.e., the end or target) is some modification in voicing, for example, *decreased strain, breathiness, and roughness*.
- Since the patient functions of *respiratory coordination during voicing*, *respiratory coordination for/during vegetative breathing*, *abdominal movement*, *rib cage movement*, and *clavicular movement* are sequentially linked, are there contexts where each of these patient functions is the criterion for a treatment's success? Yes, so they all can be unique target labels. Specifically, the two *respiratory coordination* target labels would be directly affected by different ingredients, that is, *opportunities to practice respiration in the context of voice and/or speech* versus *opportunities to practice vegetative breathing*. While respiratory coordination includes movement in the abdomen, rib cage, and clavicular area, the experts stated that there are clinical circumstances when they would individually modify

these three submovements of breathing. For example, if a patient's respiratory pattern during voicing or vegetative breathing is primarily dominated by aberrant abdominal movement patterns, one could attempt to directly and uniquely modify abdominal movement.

- Since the patient functions of *resonance* and *kinesthetic awareness* are sequentially linked, are there contexts where each of these patient functions is the criterion for a treatment's success? Yes, meaning that both target labels can be unique. The experts noted, in their clinical experience, that although *increased forward resonance* will physically increase vibrations in/around the face, the patient might not be very good at noticing these changes in kinesthetic sensations. Thus, even after resonance is modified, patients may need practice to *improve their discrimination of kinesthetic sensations*. *Improved discrimination of kinesthetic sensations* also appears to be a critical part of multiple standardized treatment protocols focused on resonant voicing (e.g., Roy et al., 2003; Verdolini-Marston et al., 1995).

The variations of Probe Question 3 produced consensus rationales that relied on attempts to identify if broad target labels (i.e., patient functions affected by numerous other patient functions) were ever directly affected by an ingredient. The following examples will concretely illustrate this principle:

- Is there a treatment that can directly affect the patient function of *pain/discomfort/soreness*? Yes, multiple treatments can directly *reduce pain/discomfort/soreness*. For example, a clinician could *apply low-level light (i.e., laser) therapy* to the anterior neck to *reduce pain* (Clijnsen et al., 2017; Cotler et al., 2015; Yousefi-Nooraie et al., 2008). Also, more concretely, *reductions in pain/discomfort/soreness* can be induced through the *provision of oral or topical analgesics*.
- Is there a treatment that can directly affect the patient function of *vocal effort*? The expert panel could not identify an ingredient that directly affected *vocal effort*; that is, ingredients only indirectly affect *vocal effort* through their hypothesized effects on other targets (e.g., *improved resonance, improved respiratory coordination during voicing*). Therefore, *changes in vocal effort* appear to always be an aim. Effort (whether defined physiologically and/or psychologically) is a measure of overall system efficiency, and treatments can target multiple functions to make the system more efficient. When a patient has attained a desirable vocal outcome (e.g., the voice sounds normal) but still reports increased vocal effort, the clinician will look for features of the patient's vocal behavior that remain suboptimal and then target those (e.g., respiration, resonance). Even if a discovery learning approach is used (e.g., "try this activity and don't use so much effort this time"; Mayer, 2004), the patient would not be expected to perform the behavior in exactly the same way but with less effort. Instead, the clinician would



expect a change in the patient's behavior (target) with resultant effort reduction (aim). Therefore, changing *vocal effort* is probably always an aim, rather than a target, because it is indirectly affected by changes in a host of different targets.

## Ingredients

During Round 4, the experts produced a similar level of agreement across the 48 ingredient labels for the concepts of measurability and uniqueness; that is, supramajority agreement was reached for 65% and 58% of ingredient labels for measurability and uniqueness/redundancy (respectively), and 38% of ingredient labels achieved supermajority consensus on both concepts. Although the vocal rehabilitation experts reported good agreement on average, there was disagreement regarding how to revise the ingredient labels to further increase uniqueness and/or decrease redundancy. Specifically, there was no consensus on which ingredient label should be chosen when multiple labels appeared redundant. To address this uncertainty, we formally operationalized the concepts of unique and redundant through the development of three ingredient probe questions in Table 3 (Van Stan et al., 2020). Ingredient labels with the most overlap among each other from Round 4 were used to create variations of the probe questions. The only revision to the list was deleting one ingredient label (*apply cold*) because all experts said they did not use that ingredient. Otherwise, there were no revisions to the ingredients list between Rounds 4 and 5.

All discussions during the in-person Round 5 meeting were scaffolded according to the probe questions. After Round 5, 40 ingredients were split, combined, or revised to attain supramajority consensus on both observability and uniqueness for two ingredients and to create 10 new ingredients.

After Round 6, 16 out of 19 ingredient labels achieved supramajority consensus on both observability and uniqueness/redundancy. Expert feedback from Round 6 resulted in revisions to the labels and definitions for the purpose of maximizing uniqueness, measurability, and general clarity. Once the ingredients were revised, the experts were asked to provide final judgments of uniqueness/redundancy and observability for the remaining three ingredients that did not attain consensus in Round 6: *provide opportunities to practice modified levels of muscle activation*, *gross vocal fold adduction exercises*, and *apply low level light therapy*. These final judgments resulted in supramajority consensus in measurability and uniqueness for all 19 ingredients and no further revisions. The next few paragraphs will describe how the probe questions guided the development of consensus on specific issues during Rounds 5 and 6.

## Ingredient Probe Questions

The variations of Probe Question 2 had the most wide-ranging consequences, so this probe question will be described first. The resulting consensus rationales were based on attempts to identify a common underlying factor among

overlapping clinician actions. Then, if an underlying commonality could be found, this commonality was identified as the unique ingredient. Subsequently, the question further probed the clinical significance of variations in this underlying common ingredient. The following examples will concretely illustrate these rationales:

- Do all manual ingredients (e.g., pressure, touch, traction, stretch) share some underlying commonality? Yes, because pressure and touch are on a continuum of how much pressure is applied, and traction, stretch, and pressure all physically occur together. Additionally, it was felt that many ingredient labels did not represent what is delivered by a clinician, but rather represent the hypothesized MoA—for example, myofascial release (Craig et al., 2015; Marszałek et al., 2012), laryngeal reposturing (Roy et al., 2017)—or a general approach/anatomical area such as massage (Laukkanen et al., 2005; Leppänen et al., 2010), laryngeal manual therapy (Mathieson et al., 2009), or circumlaryngeal massage (Roy et al., 1997; Roy & Leeper, 1993). *Provide pressure* was identified as the common denominator among all manual ingredients, because it is always present (light touch contains minimal pressure, but no traction and/or stretch) and is always directly applied by the clinician (the clinician applies pressure assuming that it will cause the sensation of touch, or create traction or stretch in the underlying tissues). Therefore, all other ingredient labels were considered redundant with the ingredient of *apply pressure*. Ingredient Probe Question 2 followed this decision with another question “What is the significance of the individual variations of [apply pressure]?” Based on this follow-up question, the variations of touch, traction, degree of pressure, and so forth were used to further operationalize the ingredient of *apply pressure* with parameters such as amount, manner, anatomical location, and so forth.
- Do the ingredients lip trill, tongue trill, lip/tongue trill, siren, inverted megaphone, and gargling all share some underlying commonality? Yes, because they all contain an SOVT. After the Round 5 discussion, it was determined that these specific SOVT techniques are different delivery vehicles that provide different dosages of the same therapeutic ingredient: resistance to phonation through increased intraoral and supraglottal pressures (Titze, 2006). Individual SOVT techniques were not further subgrouped; it is hoped that future empirical studies will guide decisions regarding when/which types of SOVT are theoretically interchangeable versus distinct ingredients. For example, incomplete lip closure may be more helpful with targeted changes in flow than complete lip closure because flow can be felt more easily when passing through a narrow opening. However, the opposite may be true in regard to targeted changes in vibrotactile sensation or resonance because acoustic pressures during voicing push more intensively against closed lips versus partially open lips.

- Do the ingredients *pitch glides*, *loudness glides*, *siren*, and *vocalises* all share an underlying commonality? Yes, because they all entail *opportunities to practice voicing*. The same could be said for ingredients that all share *opportunities to practice breathing* or *opportunities to practice sensory discrimination*. As a result, many ingredient labels were grouped into overarching common ingredient labels phrased as *provide opportunities to practice [insert function]*. Also, according to the RTSS, ingredients must be something the clinician delivers. The clinician does not deliver vocal pitch or loudness, but instead asks the patient to practice or repetitively perform a task that has a specified pitch or loudness. Sometimes, the specific pitch or loudness used during practice is meaningful for treatment and, therefore, requires specification. At other times, pitch or loudness are just inherent to voicing and not meaningful for treatment and, therefore, do not require specification. For example, *opportunities to practice sustained phonation* can have a specific amount of loudness for treatment purposes including “as loud as possible” in treatments for Lee Silverman Voice Treatment (Ramig et al., 1995, 1996), or “as soft as possible” in VFE (Roy et al., 2001, 2003; Stemple, 2005). However, if the target is *increased forward resonance*, the specific loudness of vocal practice may not be germane to the treatment (Roy et al., 2003; Verdolini-Marston et al., 1995).
- Do the ingredients with cueing and modeling all share an underlying commonality? Yes, the expert panel found two commonalities across these ingredient labels. These were grouped into two new, broader ingredient labels: *provide feedback* and *provide volition ingredient(s)*. Expert feedback indicated that splitting cueing into three different ingredients (*auditory*, *kinesthetic*, and *effort*) did not include many other concepts that could be cued (e.g., pitch, loudness, breathing, attention). It was suggested that cueing be a single category with prompts to further describe the clinically relevant ways in which the delivery of a cue may vary. This was also suggested for *provide auditory model*, as models could be provided using many more modalities (visual, physical, somatosensory, etc.). However, the various purposes of cueing or modeling were felt to most commonly fall into either *instruction on how to do a treatment activity* (a volition ingredient) or *information on performance accuracy* (feedback). For example, sometimes modeling is used as a comparator for error feedback (“you sounded like this [model] and it should sound more like this [model]”) and sometimes it is used as instructions for how to perform a desired ingredient or activity (“I want you to practice doing this: [model]”). Additionally, *provide feedback* was placed in the ingredients for direct targets category because there are well-articulated theories about how variations in feedback on performance can affect the development of skills and habits (i.e., motor learning principles; Schmidt & Lee, 2011).

The variations of Probe Question 1 produced consensus rationales that relied on determining whether multiple ingredient labels are used to affect different targets or affect the same target differently. In other words, each ingredient label can have its own unique MoA or share a common (i.e., redundant) MoA with other ingredients. The following example will illustrate this:

- Are the ingredients of *pitch glides* and *loudness glides* used interchangeably to achieve the same target(s)? It depends upon the context. There are clinical circumstances when *opportunities to practice voicing with some variation in pitch* (i.e., previously called *pitch glides*) and *opportunities to practice voicing with some variation in loudness* (i.e., previously called *loudness glides*) are unique or redundant depending upon their relationship to the treatment target. Variations in pitch and/or loudness can be redundant when added to vocal practice in the service of improved generalization. For example, *opportunities to practice voicing with natural pitch and loudness variation* hypothetically helps increase the chances that the target (e.g., *decreased vocal fry*) will generalize to spontaneous speech. However, *opportunities to practice voicing with some specific pitch or loudness* are unique ingredients when they are critical to different treatment targets. For example, asking a patient with puberphonia to practice a descending pitch glide is directly related to the target of *decreased habitual pitch* (Aronson, 1985; Roy et al., 2017). Alternatively, asking a patient with Parkinson’s disease to practice a loudness glide from softest to loudest is directly related to the target of *increased habitual loudness* (Ramig et al., 1995).

The variations of Probe Question 3 produced consensus rationales that relied on determining whether a single ingredient label that appears to affect many targets is indeed a single unique ingredient or is composed of multiple unique ingredients (each with its own MoA). The following example will illustrate this:

- Would you differentially vary the single ingredient label *phonation duration* in amount and/or manner to address different targets (e.g., *decreased strained voicing vs increased gross adduction of the true vocal folds not during voicing*)? The experts concluded that the answer is “yes.” Therefore, the ingredient can be subdivided into multiple unique ingredients based on the targets. For example, the critical aspects of repetitious movement (i.e., phonation duration) are different if one wants to acquire a skill/habit (i.e., *decreased strained voicing*) versus increased organ function (i.e., *increased gross adduction of the true vocal folds*): Opportunities to practice sustained voicing in a specific way (probably combined with *providing feedback*) versus *providing resistance to muscle contraction during voicing*, respectively. Of note, while the latter ingredient is hypothetically possible, it is not in the final ingredients list. This is because all experts stated that

they do not measure movement, strength, or endurance of the laryngeal muscles/structures to evaluate the success of any phonatory duration-related ingredient.

## Discussion

This study resulted in the creation of a new standardized tool called the RTSS-Voice that can organize and facilitate communication surrounding the systematic study of vocal rehabilitation treatments according to the RTSS framework. Specifically, the RTSS-Voice contains expert consensus labels for 35 target categories and 19 ingredient categories with operational definitions outlining how targets or ingredients within each category may vary. Multiple types of validity have been established for the RTSS-Voice including face validity (categories are measurable and unique), content validity (categories represent many constituents of treatment in vocal rehabilitation), and construct validity (categories accurately reflect the rules and terminology of the RTSS). Additionally, the project demonstrated that it is possible to reach consensus among multiple clinicians about the critical ingredients and targets from a field of treatment, as consensus was not an assumed outcome. Every clinical expert began the Delphi process with differing treatment theories, training backgrounds, experiences, personalized terminology to represent their ideas, and vested interests in specific approaches to treatment. Using the RTSS as a common framework for discussion was likely a major contributor to the success of this endeavor. The RTSS essentially required the expert clinicians to translate their individual perspectives into a shared theoretical and conceptual language. Intractable differences often became tractable after translating clinical viewpoints into the RTSS, for example, see the previously described examples in response to various probe questions.

The RTSS-Voice may represent three significant theoretical and practical advances to current voice therapy practice and clinical reasoning. First, its treatment ingredients focus on the clinician's actions in reference to the treatment target (e.g., *provide opportunities to practice voicing in a specific manner*) instead of simply what the patient does (e.g., *pitch glides, half-swallow boom*) or the target of the treatment (e.g., *resonant voice, flow phonation*). This is of value because the former explicitly tells the clinician what to do to deliver the treatment and directs the clinician's attention to what is therapeutically important about the ingredient (repetitive voicing is practice for building a skill or habit, not to build muscle strength, endurance, range of motion, etc.), while the latter hints at neither. Second, ingredients explicitly map to their individual target, instead of the current practice where ingredients and targets are simply listed in separate categories. As stated in the introduction, these explicit connections are necessary to understand and systematically refine/define voice therapy because clinicians often provide multiple ingredients for different purposes (targets). Third, the patient functions thought to be directly (targets) and indirectly (aims) modified by the treatment ingredients are explicitly distinguished in a causal

manner, whereas current practice simply lists modified patient functions as primary or secondary outcomes based on their level of clinical importance and secondary effects. Identifying primary and secondary outcomes without aligning them to the concepts of targets and aims (i.e., without putative causal links between the outcomes and ingredients provided) can be problematic. For example, the primary outcome for many voice treatment approaches is a broad measure such as the Voice-Related Quality of Life (Hogikyan & Sethuraman, 1999), the VHI (Jacobson et al., 1997), or a perceptual or objective estimate of overall dysphonia. However, according to the RTSS-Voice, quality of life, handicap, and overall dysphonia are probably always the aim for treatment and require the successful modification of multiple targets. In other words, the ingredients have no direct effects on the primary outcome (aim) and, therefore, each target still needs measurement. Thus, to understand how or which ingredients contribute to changes (or no changes) in the primary outcome, some of the secondary outcomes will need to be theoretically designated as the target for an ingredient(s) (hopefully, a priori). However, secondary outcomes often consist of a standard battery of acoustic, aerodynamic, and endoscopic measures that are compulsorily acquired (Dejonckere et al., 2001; Patel et al., 2018); that is, the measures may not always be chosen for their explicit connection to various treatment ingredients provided. In addition, if any treatment targets are not aligned with a secondary outcome in the standard clinical assessment, they will not have been measured at all. Or worse, targets could be associated with secondary outcomes a posteriori according to whichever traditionally acquired secondary measure significantly changed, without any theory-driven rationale.

The development of standard operational definitions and labels is a significant step forward because they provide a theory-driven lexicon and rules that can enhance the clarity of conceptual thinking and communication among voice researchers, clinicians, and educators. For example, after multiple voice treatment protocols are specified in a standard manner according to the RTSS-Voice, researchers could undertake comparative effectiveness research and meta-analyses based on known differences and similarities in terms of ingredients, dosage of ingredients, and targets. This would be a marked improvement over current meta-analysis and comparative effectiveness approaches that are limited to comparing entire protocols (Van Stan et al., 2019), including underspecified standard or usual care, a very common comparator in treatment research (Lohse et al., 2018; Whyte et al., 2018). If frontline clinicians use the RTSS-Voice to describe their standard care, researchers could design and test standardized treatment protocols that intentionally do or do not contain ingredients and targets that are commonly used in standard care. In addition, frontline clinicians could more easily recognize the differences between their treatment and a new or existing research treatment protocol (i.e., improved implementation and dissemination of evidence-based practice). Finally, if researchers and clinicians widely adopt the RTSS-Voice, educators who train clinicians and students could teach treatment concepts in



a way that should directly generalize to understanding/ applying the research literature and the clinical practice of clinician mentors.

Future work is needed to improve, supplement, and/or streamline the RTSS-Voice based on attempts to describe research treatments (e.g., individual protocols and the similarities/differences among protocols) and treatment provided in everyday standard care (e.g., what ingredients and targets are being used frequently, differently, or not at all by frontline clinicians) using a standardized approach. Empirical work such as this could result in an online toolkit that helps guide researchers or clinicians to specify their treatment protocols or everyday care, perhaps eventually leading to the development of an electronic medical record based on the RTSS-Voice. In addition to increased clarity regarding the hypothesized active elements of treatment, these categories have great potential to broadly improve the implementation of the RTSS in vocal rehabilitation treatment research, education, and everyday clinical care. For example, individuals would not have to “start from scratch” to describe a voice treatment with sufficient clarity to be replicable. Additionally, the methods used in this project may help others to apply the RTSS outside the field of voice therapy for two reasons. First, the methods and results are generalizable across rehabilitation because they are based on the broad RTSS framework. Second, the project provides a concrete input example (a current practice product: the taxonomy of voice therapy) and output example (the RTSS-Voice) of the RTSS specification process. Ultimately, once the RTSS-Voice is used in research protocols and standard care, the theoretical and clinical premises underlying the creation of the RTSS-Voice can be explicitly tested. These include questions about the construct validity of the RTSS such as the following: (a) Compared to existing research treatment descriptions, do RTSS-based specifications improve treatment fidelity, efficacy, comparative effectiveness, or meta-analyses? (b) Compared to current approaches to thinking about treatment, does using the RTSS result in improved clinical reasoning? (c) Compared to current standard practices, do patient outcomes improve when clinicians use the RTSS to frame the treatment provided? Additionally, broad use of the RTSS enables new testable questions about treatment principles that are broader than a single protocol or subfield of rehabilitation, such as the following: (a) What are the volition ingredients and methods of tailoring them that maximize performance of a wide range of treatment activities taking place outside of the clinic? (b) How does the amount and schedule of practice affect overall voice quality, and does it depend upon the specific vocal behavior being practiced?

### Study Limitations

For the most part, this study did not include targets and ingredients associated with two groups of treatment components: *Activity-like Skills and Habits* and *Representations*. Although these treatment components were excluded due to the potentially infinite number of targets and ingredients (there are countless different skills and informational

topics), the results of this study suggest these components might be composed of systematic categories and underlying ways in which they can vary. That is, the methodology outlined here may be adapted to identify common classes of ingredients and targets that cover many different skilled activities and mental states. For example, an ingredient category for *Representations* treatment components could be *Provide information on [insert topic]* with variations in the *Information* delivered (e.g., vocal hygiene recommendations), *Delivery method* (e.g., verbal, written), *Delivery vehicle* (e.g., analogy, list of points), *Difficulty* (e.g., level of abstraction, depth of information), and *Dose* (e.g., bits of information per time or phrase). Also, an ingredient category for *Activity-like Skills and Habits* treatment components could be *Provide opportunities to practice [insert activity]* with variations in the *Practice structure* (e.g., part versus whole practice), *Variability* (e.g., different situations when the activity will be performed), *Difficulty* (e.g., speech material, situational factors, emotional content), and *Dose* (e.g., amount of practice to achieve a level of mastery).

The RTSS-Voice does not comprehensively capture treatments involving nonlaryngeal sound sources (e.g., tracheoesophageal prostheses, electro-larynges) or situations where external aides enable a laryngeal sound source (e.g., tracheostomy tubes and associated speaking valves). Therefore, the RTSS-Voice does not include these ingredients; future work will be necessary to incorporate them. However, the RTSS-Voice may cover most ingredients and targets for patients without a larynx or a tracheostomy, as treatment is likely composed of components that are common across voice therapy regardless of the disorder. Specifically, a patient using a tracheoesophageal prosthesis, electrolarynx, or speaking valve will probably need opportunities to practice the skill and/or habit of voicing with these devices while the clinician provides feedback.

Finally, the results of a Delphi-based methodology reflect the expertise, backgrounds, and biases of those who participate. However, we designed rigorous methods to identify and neutralize bias during the Delphi process (i.e., a facilitator and two external readers with no expertise in vocal rehabilitation) and ensure the highest quality and diversity of expert opinion (i.e., internationally recognized vocal rehabilitation experts with varied expertise from clinical research labs and multiple voice centers). Additionally, while the responses for each expert were de-identified during the Delphi Rounds (except for the in-person Delphi Round 2), the experts knew who else was generally involved in the process. It was not possible to maintain strict anonymity of who was involved in the study for multiple practical reasons, for example, experts knowing each other from previous collaborations, meeting at conferences, or working together at institutions before or after the study began.

### RTSS-Voice Implementation

The final treatment targets and ingredients contained in the appendixes may seem to contain an overwhelming amount of detail for practical implementation. However,

**Table 4.** Case Study 1 narrative and Rehabilitation Treatment Specification System treatment specification.

**Narrative:**

A patient with unilateral vocal fold paralysis is 1-week status post a surgical medialization and still voices with severely strained voice quality. The aberrant voice quality appears to have a large behavioral component, because the patient's strained voice quality significantly decreases when he is asked to voice in a specific way: sustained voicing for 3–5 s with forward resonance during nasal semi-occluded vocal tract (SOVT) postures (e.g., /m, n, ŋ/). Therefore, the patient is asked to practice voicing with forward resonance (ingredient) to decrease his strained voice quality (target). Although the patient must breathe, activate certain levels of muscle contraction, and perceive the sensory consequences of phonation during vocal practice, there is no requirement to specify ingredients related to practicing breathing, muscle activation during voicing, or sensory discrimination. This is because these are not the hypothesized ingredients or targets of the treatment (according to this clinician's treatment theory for this specific patient). Also, the clinician provides feedback on the patient's performance starting at 100% frequency (after every practice trial) and fades to no feedback as the patient maintains 80% accuracy (with a hypothesis that this feedback schedule will help solidify new vocal motor patterns).

Further specification of the ingredient "opportunities to practice voicing" depends upon what aspects of the ingredient are relevant to the clinician's treatment theory. In this case, to decrease strained voice quality, the clinician wants the patient to practice increased forward resonance during voicing (which can reduce strain), while progressing through a difficulty hierarchy of sustained phonation (3–5 s) first and then speech (sustained vowels are easier for patient success than spontaneous speech and can be quickly generalized to speech), and using speech-based SOVTs like /o, u, m, n, ŋ, z/. The SOVT postures make it physiologically easier to produce and feel forward resonance. Also, speech-based SOVTs are thought to better facilitate generalization of forward resonance into speech than using external devices like straws, kazoo's, and so forth. Although voicing inherently has many other aspects (e.g., loudness, pitch, mean airflow), they are not germane to the therapist's treatment theory and, therefore, do not need to be specified.

Since the treatment described is volitional in nature, the clinician must think about how the ingredients are oriented toward increasing the patient's Capability, Opportunity, and Motivation to perform the desired Behavior (COM-B). How does the patient know what he should practice (Capability)? In this case, "how to" instruction is provided by a clinician vocal model, which demonstrates the desired voicing features. To help Motivate the patient, the clinician provides a rationale for why the ingredients and target are being used. Because the patient is practicing in front of the clinician, no additional ingredients are provided to increase the patient's Opportunity to perform the behavior.

Target	Ingredients	Dose
Decreased strain	Opportunities to practice voicing	Practice
Target measure:	(a) forward resonance	No. of repetitions until 80% accurate without feedback.
Clinician's perceptual rating on a 100-mm visual analog scale	(b) speech-based SOVTs like /o, u, m, n, ŋ, z/ (c) Difficulty hierarchy beginning with sustained phonation (3–5 s) and transitioning to speech	Progression rule: increase difficulty when patient produces minimal strain ~100% accuracy
	Provide feedback	Feedback
	(a) Clinician-delivered	Start at 100% frequency and fade by 25% frequency increments when 80% accurate for 30 trials
	(b) Verbal	
	(c) Knowledge of results (after practice trial)	
	Provide volition ingredients	Volition
	(1) Provide instructions: template for voicing practice as vocally modeled by the clinician	As needed until patient produces correct behavior
	(2) Provide rationale for ingredients and target	As needed until patient appears motivated

in research, clinical, or educational settings, the clinician would not need to document every potential way in which the targets and ingredients could have varied, only those that are theoretically meaningful to the specific treatment. In other words, a clinician would only specify information about a target or ingredient when that information is hypothetically important for achieving a therapeutic effect. For example, if an ingredient is selected simply because of clinician preference or comfort (e.g., using a straw for an SOVT instead of another delivery vehicle like a kazoo or /m/), it is only necessary to specify the use of an SOVT (not the straw or any specific characteristics of the straw such as diameter, length, etc.). However, if the SOVT delivery vehicle is chosen because of the level of resistance it provides, then either the specific vehicle or its resistance properties would be specified. Expecting a clinical researcher to meticulously define and report a treatment under investigation does not seem unreasonable. Once such a treatment's impact is clearly established, clinicians implementing it may refer to it by its name, while still being able to define and, when necessary, audit its active ingredients. Additionally, while considerable time would probably be required for frontline clinicians to

initially specify the treatment(s) provided, the resulting specification template can likely be used in a time-efficient manner to document subsequent therapy. Once the clinician has a reusable specification template, the clinician would only need to fill in the template according to how the treatment was individually tailored to the patient. To demonstrate the use of the RTSS-Voice, Table 4 provides a hypothetical case study with example specification and rationales.

## Conclusions

The RTSS-based Delphi process resulted in a comprehensive and standard description of vocal rehabilitation treatment called the RTSS-Voice. Specifically, the RTSS-Voice consists of named and operationalized categories of measurable targets and commonly used ingredients known or hypothesized to affect those targets. With this set, in theory, every current voice therapy approach could be specified in comparable terms. Future treatment developers could explicitly assess whether they are contributing unique ingredients to the treatment armamentarium or novel ingredient combinations. The RTSS-Voice has potential to improve



voice therapy education/training, clinical care, and research. Additionally, the methods developed here could be useful in other subspecialty areas of speech-language pathology (e.g., motor speech disorders, aphasia, dysphagia) and rehabilitation in general (e.g., physical therapy, occupational therapy, nursing). Further refinements of the RTSS-Voice, as well as its adoption to specify voice therapy protocols and standard care, will require collaboration among creators of the treatment protocols, frontline clinicians, and experts in the RTSS framework. The adoption of the RTSS has great potential to help transition the evidence base of voice treatment from its current state (i.e., entire treatment protocols work for some patients for unknown reasons) to an understanding of which clinician action(s) affect specific patient functions.

## Acknowledgments

This work was supported by the National Institute on Deafness and Other Communication Disorders and the Office of Behavioral and Social Sciences Research via Grant R21 DC016124 (PI: Jarrad H. Van Stan). The article's contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

## References

- Altman, K. W., Atkinson, C., & Lazarus, C. (2005). Current and emerging concepts in muscle tension dysphonia: A 30-month review. *Journal of Voice, 19*(2), 261–267. <https://doi.org/10.1016/j.jvoice.2004.03.007>
- Aronson, A. (1985). *Clinical voice disorders* (2nd ed.). Thieme.
- Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America, 119*(4), 2363–2371. <https://doi.org/10.1121/1.2173513>
- Behlau, M., & Oliveira, G. (2009). Vocal hygiene for the voice professional. *Current Opinion in Otolaryngology & Head & Neck Surgery, 17*(3), 149–154. <https://doi.org/10.1097/MOO.0b013e32832af105>
- Behrman, A. (2006). Facilitating behavioral change in voice therapy: The relevance of motivational interviewing. *American Journal of Speech-Language Pathology, 15*(3), 215–225. [https://doi.org/10.1044/1058-0360\(2006\)020](https://doi.org/10.1044/1058-0360(2006)020)
- Behrman, A., Rutledge, J., Hembree, A., & Sheridan, S. (2008). Vocal hygiene education, voice production therapy, and the role of patient adherence: A treatment effectiveness study in women with phonotrauma. *Journal of Speech, Language, and Hearing Research, 51*(2), 350–366. [https://doi.org/10.1044/1092-4388\(2008\)026](https://doi.org/10.1044/1092-4388(2008)026)
- Burnett, T. A., & Larson, C. R. (2002). Early pitch-shift response is active in both steady and dynamic voice pitch control. *The Journal of the Acoustical Society of America, 112*(3), 1058–1063. <https://doi.org/10.1121/1.1487844>
- Clijse, R., Brunner, A., Barbero, M., Clarys, P., & Taeymans, J. (2017). Effects of low-level laser therapy on pain in patients with musculoskeletal disorders—A systematic review and meta-analysis. *European Journal of Physical and Rehabilitation Medicine, 53*(4), 603–610. <https://doi.org/10.23736/S1973-9087.17.04432-X>
- Cotler, H. B., Chow, R. T., Hamblin, M. R., & Carroll, J. (2015). The use of low level laser therapy (LLLT) for musculoskeletal pain. *MOJ Orthopedics & Rheumatology, 2*(5), 00068. <https://doi.org/10.15406/mojor.2015.02.00068>
- Craig, J., Tomlinson, C., Stevens, K., Kotagal, K., Fornadley, J., Jacobson, B., Garrett, C. G., & Francis, D. O. (2015). Combining voice therapy and physical therapy: A novel approach to treating muscle tension dysphonia. *Journal of Communication Disorders, 58*, 169–178. <https://doi.org/10.1016/j.jcomdis.2015.05.001>
- Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., Van De Heyning, P., Remacle, M., & Woisard, V. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *European Archives of Oto-Rhino-Laryngology, 258*(2), 77–82. <https://doi.org/10.1007/s004050000299>
- Dijkers, M. P., Kropp, G. C., Esper, R. M., Yavuzer, G., Cullen, N., & Bakdali, Y. (2002). Quality of intervention research reporting in medical rehabilitation journals. *American Journal of Physical Medicine & Rehabilitation, 81*(1), 21–33. <https://doi.org/10.1097/00002060-200201000-00005>
- Emerson, J. D., Burdick, E., Hoaglin, D. C., Mosteller, F., & Chalmers, T. C. (1990). An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Controlled Clinical Trials, 11*(5), 339–352. [https://doi.org/10.1016/0197-2456\(90\)90175-2](https://doi.org/10.1016/0197-2456(90)90175-2)
- Gartner-Schmidt, J., Gherson, S., Hapner, E. R., Muckala, J., Roth, D., Schneider, S., & Gillespie, A. I. (2016). The development of conversation training therapy: A concept paper. *Journal of Voice, 30*(5), 563–573. <https://doi.org/10.1016/j.jvoice.2015.06.007>
- Gillespie, A. I., Gartner-Schmidt, J., Rubinstein, E. N., & Verdolini Abbott, K. (2013). Aerodynamic profiles of women with muscle tension dysphonia/aphonia. *Journal of Speech, Language, and Hearing Research, 56*(2), 481–488. [https://doi.org/10.1044/1092-4388\(2012\)11-0217](https://doi.org/10.1044/1092-4388(2012)11-0217)
- Gillespie, A. I., Yabes, J., Rosen, C. A., & Gartner-Schmidt, J. L. (2019). Efficacy of conversation training therapy for patients with benign vocal fold lesions and muscle tension dysphonia compared to historical matched control patients. *Journal of Speech, Language, and Hearing Research, 62*(11), 4062–4079. [https://doi.org/10.1044/2019\\_JSLHR-S-19-0136](https://doi.org/10.1044/2019_JSLHR-S-19-0136)
- Gilman, M., Maira, C., & Hapner, E. R. (2019). Airflow patterns of running speech in patients with voice disorders. *Journal of Voice, 33*(3), 277–283. <https://doi.org/10.1016/j.jvoice.2017.12.004>
- Hancock, A. B., & Garabedian, L. M. (2013). Transgender voice and communication treatment: A retrospective chart review of 25 cases. *International Journal of Language & Communication Disorders, 48*(1), 54–65. <https://doi.org/10.1111/j.1460-6984.2012.00185.x>
- Hart, T., Dijkers, M. P., Whyte, J., Turkstra, L. S., Zanca, J. M., Packel, A., Van Stan, J. H., Ferraro, M., & Chen, C. (2019). A theory-driven system for the specification of rehabilitation treatments. *Archives of Physical Medicine and Rehabilitation, 100*(1), 172–180. <https://doi.org/10.1016/j.apmr.2018.09.109>
- Hart, T., Tsoulosides, T., Zanca, J. M., Whyte, J., Packel, A., Ferraro, M., & Dijkers, M. P. (2014). Toward a theory-driven classification of rehabilitation treatments. *Archives of Physical Medicine and Rehabilitation, 95*(1), S33–S44. e32. <https://doi.org/10.1016/j.apmr.2013.05.032>
- Hillman, R. E., Stepp, C., Van Stan, J. H., Zanartu, M., & Mehta, D. D. (2020). An updated theoretical framework for vocal hyperfunction. *American Journal of Speech-Language Pathology, 29*(4), 2254–2260. [https://doi.org/10.1044/2020\\_AJSLP-20-00104](https://doi.org/10.1044/2020_AJSLP-20-00104)

- Hogikyan, N. D., & Sethuraman, G. (1999). Validation of an instrument to measure Voice-Related Quality Of Life (V-RQOL). *Journal of Voice*, 13(4), 557–569. [https://doi.org/10.1016/S0892-1997\(99\)80010-1](https://doi.org/10.1016/S0892-1997(99)80010-1)
- Holmberg, E. B., Hillman, R. E., Hammarberg, B., Sodersten, M., & Doyle, P. (2001). Efficacy of a behaviorally based voice therapy protocol for vocal nodules. *Journal of Voice*, 15(3), 395–412. [https://doi.org/10.1016/S0892-1997\(01\)00041-8](https://doi.org/10.1016/S0892-1997(01)00041-8)
- Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The Voice Handicap Index (VHI). *American Journal of Speech-Language Pathology*, 6(3), 66–70. <https://doi.org/10.1044/1058-0360.0603.66>
- Kapsner-Smith, M. R., Hunter, E. J., Kirkham, K., Cox, K., & Titze, I. R. (2015). A randomized controlled trial of two semi-occluded vocal tract voice therapy protocols. *Journal of Speech, Language, and Hearing Research*, 58(3), 535–549. [https://doi.org/10.1044/2015\\_JSLHR-S-13-0231](https://doi.org/10.1044/2015_JSLHR-S-13-0231)
- Kochis-Jennings, K. A., Finnegan, E. M., Hoffman, H. T., Jaiswal, S., & Hull, D. (2014). Cricothyroid muscle and thyroarytenoid muscle dominance in vocal register control: preliminary results. *Journal of Voice*, 28(4), 652. e621–652. e629. <https://doi.org/10.1016/j.jvoice.2014.01.017>
- Laukkanen, A. M., Leppänen, K., Tyrmi, J., & Vilkmán, E. (2005). Immediate effects of ‘Voice Massage’ treatment on the speaking voice of healthy subjects. *Folia Phoniatrica et Logopaedica*, 57(3), 163–172. <https://doi.org/10.1159/000084136>
- Leppänen, K., Ilomäki, I., & Laukkanen, A. M. (2010). One-year follow-up study of self-evaluated effects of Voice Massage™, voice training, and voice hygiene lecture in female teachers. *Logopedics Phoniatrics Vocology*, 35(1), 13–18. <https://doi.org/10.3109/14015430903552360>
- Linstone, H. A., & Turoff, M. (1975). *The Delphi method*. Addison-Wesley Reading, MA.
- Lohse, K. R., Pathania, A., Wegman, R., Boyd, L. A., & Lang, C. E. (2018). On the reporting of experimental and control therapies in stroke rehabilitation trials: A systematic review. *Archives of Physical Medicine and Rehabilitation*, 99(7), 1424–1432. <https://doi.org/10.1016/j.apmr.2017.12.024>
- Lowell, S. Y., & Story, B. H. (2006). Simulated effects of cricothyroid and thyroarytenoid muscle activation on adult-male vocal fold vibration. *The Journal of the Acoustical Society of America*, 120(1), 386–397. <https://doi.org/10.1121/1.2204442>
- Marszałek, S., Niebudek-Bogusz, E., Woźnicka, E., Malińska, J., Golusiński, W., & Śliwińska-Kowalska, M. (2012). Assessment of the influence of osteopathic myofascial techniques on normalization of the vocal tract functions in patients with occupational dysphonia. *International Journal of Occupational Medicine and Environmental Health*, 25(3), 225–235. <https://doi.org/10.2478/S13382-012-0041-7>
- Mathieson, L., Hirani, S. P., Epstein, R., Baken, R. J., Wood, G., & Rubin, J. S. (2009). Laryngeal manual therapy: A preliminary study to examine its treatment effects in the management of muscle tension dysphonia. *Journal of Voice*, 23(3), 353–366. <https://doi.org/10.1016/j.jvoice.2007.10.002>
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, 59(1), 14–19. <https://doi.org/10.1037/0003-066X.59.1.14>
- McNeill, E. J. (2006). Management of the transgender voice. *The Journal of Laryngology & Otolaryngology*, 120(7), 521–523. <https://doi.org/10.1017/S0022215106001174>
- Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, 46(1), 81–95. <https://doi.org/10.1007/s12160-013-9486-6>
- Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(1), 42. <https://doi.org/10.1186/1748-5908-6-42>
- Morrison, M. D., Nichol, H., & Rammage, L. A. (1986). Diagnostic criteria in functional dysphonia. *The Laryngoscope*, 96(1), 1–8. <https://doi.org/10.1288/00005537-198601000-00001>
- Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology*, 27(3), 887–905. [https://doi.org/10.1044/2018\\_AJSLP-17-0009](https://doi.org/10.1044/2018_AJSLP-17-0009)
- Patel, R. R., Dubrovskiy, D., & Döllinger, M. (2014). Characterizing vibratory kinematics in children and adults with high-speed digital imaging. *Journal of Speech, Language, and Hearing Research*, 57(2), S674–S686. [https://doi.org/10.1044/2014\\_JSLHR-S-12-0278](https://doi.org/10.1044/2014_JSLHR-S-12-0278)
- Pitts, T., Bolser, D., Rosenbek, J., Troche, M., Okun, M. S., & Sapienza, C. M. (2009). Impact of expiratory muscle strength training on voluntary cough and swallow function in Parkinson disease. *Chest*, 135(5), 1301–1308. <https://doi.org/10.1378/chest.08-1389>
- Ramig, L. O., Countryman, S., O’Brien, C., Hoehn, M., & Thompson, L. (1996). Intensive speech treatment for patients with Parkinson’s disease: Short- and long-term comparison of two techniques. *Neurology*, 47(6), 1496–1504. <https://doi.org/10.1212/WNL.47.6.1496>
- Ramig, L. O., Countryman, S., Thompson, L. L., & Horii, Y. (1995). Comparison of two forms of intensive speech treatment for Parkinson disease. *Journal of Speech and Hearing Research*, 38(6), 1232–1251. <https://doi.org/10.1044/jshr.3806.1232>
- Roy, N., Bless, D. M., Heisey, D., & Ford, C. N. (1997). Manual circumlaryngeal therapy for functional dysphonia: An evaluation of short- and long-term treatment outcomes. *Journal of Voice*, 11(3), 321–331. [https://doi.org/10.1016/S0892-1997\(97\)80011-2](https://doi.org/10.1016/S0892-1997(97)80011-2)
- Roy, N., Gray, S. D., Simon, M., Dove, H., Corbin-Lewis, K., & Stemple, J. C. (2001). An evaluation of the effects of two treatment approaches for teachers with voice Disorders. *Journal of Speech, Language, and Hearing Research*, 44(2), 286–296. [https://doi.org/10.1044/1092-4388\(2001\)023](https://doi.org/10.1044/1092-4388(2001)023)
- Roy, N., & Leeper, H. A. (1993). Effects of the manual laryngeal musculoskeletal tension reduction technique as a treatment for functional voice disorders: Perceptual and acoustic measures. *Journal of Voice*, 7(3), 242–249. [https://doi.org/10.1016/S0892-1997\(05\)80333-9](https://doi.org/10.1016/S0892-1997(05)80333-9)
- Roy, N., Peterson, E. A., Pierce, J. L., Smith, M. E., & Houtz, D. R. (2017). Manual laryngeal reposturing as a primary approach for mutational falsetto. *The Laryngoscope*, 127(3), 645–650. <https://doi.org/10.1002/lary.26053>
- Roy, N., Weinrich, B., Gray, S. D., Tanner, K., Stemple, J. C., & Sapienza, C. M. (2003). Three treatments for teachers with voice disorders. *Journal of Speech, Language, and Hearing Research*, 46(3), 670–688. [https://doi.org/10.1044/1092-4388\(2003\)053](https://doi.org/10.1044/1092-4388(2003)053)
- Roy, N., Weinrich, B., Gray, S. D., Tanner, K., Toledo, S. W., Dove, H., Corbin-Lewis, K., & Stemple, J. C. (2002). Voice amplification versus vocal hygiene instruction for teachers with voice disorders. *Journal of Speech, Language, and Hearing Research*, 45(4), 625–638. [https://doi.org/10.1044/1092-4388\(2002\)050](https://doi.org/10.1044/1092-4388(2002)050)

- Schmidt, R. A., & Lee, T. D. (2011). *Motor control and learning: A behavioral emphasis* (Vol. 5th). Human Kinetics.
- Stemple, J. C. (2005). A holistic approach to voice therapy. *Seminars in Speech and Language, 26*(2), 131–137. <https://doi.org/10.1055/s-2005-871209>
- Story, B. H., Laukkanen, A.-M., & Titze, I. R. (2000). Acoustic impedance of an artificially lengthened and constricted vocal tract. *Journal of Voice, 14*(4), 455–469. [https://doi.org/10.1016/S0892-1997\(00\)80003-X](https://doi.org/10.1016/S0892-1997(00)80003-X)
- Stathopoulos, E. T., Huber, J. E., Richardson, K., Kamphaus, J., DeCicco, D., Darling, M., Fulcher, K., & Sussman, J. E. (2014). Increased vocal intensity due to the Lombard effect in speakers with Parkinson's disease: Simultaneous laryngeal and respiratory strategies. *Journal of Communication Disorders, 48*(March-April), 1–17. <https://doi.org/10.1016/j.jcomdis.2013.12.001>
- Titze, I. R. (2006). Voice training and therapy with a semi-occluded vocal tract: Rationale and scientific underpinnings. *Journal of Speech, Language, and Hearing Research, 49*(2), 448–459. [https://doi.org/10.1044/1092-4388\(2006\)035](https://doi.org/10.1044/1092-4388(2006)035)
- Turkstra, L. S., Norman, R., Whyte, J., Dijkers, M. P., & Hart, T. (2016). Knowing what we're doing: Why specification of treatment methods is critical for evidence-based practice in speech-language pathology. *American Journal of Speech-Language Pathology, 25*(2), 164–171. [https://doi.org/10.1044/2015\\_AJSLP-15-0060](https://doi.org/10.1044/2015_AJSLP-15-0060)
- Van Houtte, E., Van Lierde, K., & Claeys, S. (2011). Pathophysiology and treatment of muscle tension dysphonia: A review of the current knowledge. *Journal of Voice, 25*(2), 202–207. <https://doi.org/10.1016/j.jvoice.2009.10.009>
- van Leer, E., & Connor, N. P. (2012). Use of portable digital media players increases patient motivation and practice in voice therapy. *Journal of Voice, 26*(4), 447–453. <https://doi.org/10.1016/j.jvoice.2011.05.006>
- Van Stan, J. H., Dijkers, M. P., Whyte, J., Hart, T., Turkstra, L. S., Zanca, J. M., & Chen, C. (2019). The Rehabilitation Treatment Specification System: Implications for improvements in research design, reporting, replication, and synthesis. *Archives of Physical Medicine and Rehabilitation, 100*(1), 146–155. <https://doi.org/10.1016/j.apmr.2018.09.112>
- Van Stan, J. H., Roy, N., Awan, S., Stemple, J., & Hillman, R. E. (2015). A taxonomy of voice therapy. *American Journal of Speech-Language Pathology, 24*(2), 101–125. [https://doi.org/10.1044/2015\\_AJSLP-14-0030](https://doi.org/10.1044/2015_AJSLP-14-0030)
- Van Stan, J. H., Whyte, J., Duffy, J. R., Barkmeier-Kraemer, J. M., Doyle, P. B., Gherson, S., Kelchner, L., Muise, J., Petty, B., Roy, N., Stemple, J., Thibeault, S., & Jorgensen Tolejano, C. (2020). Rehabilitation Treatment Specification System: Methodology to identify and describe unique targets and ingredients. *Archives of Physical Medicine & Rehabilitation, 102*(3), 521–531. <https://doi.org/10.1016/j.apmr.2020.09.383>
- Verdolini-Marston, K., Burke, M. K., Lessac, A., Glaze, L., & Caldwell, E. (1995). Preliminary study of two methods of treatment for laryngeal nodules. *Journal of Voice, 9*(1), 74–85. [https://doi.org/10.1016/S0892-1997\(05\)80225-5](https://doi.org/10.1016/S0892-1997(05)80225-5)
- Watts, C. R., Hamilton, A., Toles, L. E., Childs, L., & Mau, T. (2015). A randomized controlled trial of stretch-and-flow voice therapy for muscle tension dysphonia. *The Laryngoscope, 125*(6), 1420–1425. <https://doi.org/10.1002/lary.25155>
- Watts, C. R., Hamilton, A., Toles, L. E., Childs, L., & Mau, T. (2019). Intervention outcomes of two treatments for muscle tension dysphonia: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research, 62*(2), 272–282. [https://doi.org/10.1044/2018\\_JSLHR-S-18-0118](https://doi.org/10.1044/2018_JSLHR-S-18-0118)
- Whyte, J., Dijkers, M. P., Hart, T., Van Stan, J. H., Packel, A., Turkstra, L. S., Zanca, J. M., Chen, C., & Ferraro, M. (2019). The importance of voluntary behavior in rehabilitation treatment and outcomes. *Archives of Physical Medicine and Rehabilitation, 100*(1), 156–163. <https://doi.org/10.1016/j.apmr.2018.09.111>
- Whyte, J., Dijkers, M. P., Hart, T., Zanca, J. M., Packel, A., Ferraro, M., & Tsaousides, T. (2014). Development of a theory-driven rehabilitation treatment taxonomy: Conceptual issues. *Archives of Physical Medicine and Rehabilitation, 95*(1), S24–S32.e22. <https://doi.org/10.1016/j.apmr.2013.05.034>
- Whyte, J., Dijkers, M. P., Van Stan, J. H., & Hart, T. (2018). Specifying what we study and implement in rehabilitation: Comments on the reporting of clinical research. *Archives of Physical Medicine and Rehabilitation, 99*(7), 1433–1435. <https://doi.org/10.1016/j.apmr.2018.03.008>
- Whyte, J., & Hart, T. (2003). It's more than a black box; It's a Russian doll. *American Journal of Physical Medicine & Rehabilitation, 82*(8), 639–652. <https://doi.org/10.1097/01.PHM.0000078200.61840.2D>
- Wood, C. E., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., & Michie, S. (2015). Applying the behaviour change technique (BCT) taxonomy v1: A study of coder training. *Translational Behavioral Medicine, 5*(2), 134–148. <https://doi.org/10.1007/s13142-014-0290-z>
- World Health Organization. (2001). *International classification of functioning, disability, and health*.
- Yousefi-Nooraie, R., Schonstein, E., Heidari, K., Rashidian, A., Pennick, V., Akbari-Kamrani, M., Irani, S., Shakiba, B., Hejri, S. M., & Jonaidi, A. (2008). Low level laser therapy for non-specific low-back pain. *Cochrane Database of Systematic Reviews, 16*(2), CD005107. <https://doi.org/10.1002/14651858.CD005107.pub4>
- Ziegler, A., Verdolini Abbott, K., Johns, M., Klein, A., & Hapner, E. R. (2014). Preliminary data on two voice therapy interventions in the treatment of presbyphonia. *The Laryngoscope, 124*(8), 1869–1876. <https://doi.org/10.1002/lary.24548>



---

The Use of “Increase” and/or “Decrease” With Skills and Habits Targets

---

Some Skills and Habits targets have “increase” and/or “decrease” under “Change in what way” because there are clinical situations in which simply increasing or decreasing a function is desired (not just in terms of “accuracy”); that is, the function cannot be thought of as increasing or decreasing too much. For example:

- A patient with bilateral vocal fold nodules may be given ingredients to “increase pitch range.”
- A patient with Parkinson’s disease may be given ingredients to “increase loudness.”
- A patient with a unilateral vocal fold sulcus may be given ingredients to “decrease roughness.”

---

The Use of Clinician or Patient Judgment for Measuring Targets

---

When “clinician judgment” or “patient judgment” is/are listed under “measurement” of a target, this can include informal perceptual scales. However, formalized perceptual scales with established estimates of reliability and validity would be preferable.

---

Targets

---

***The list of consensus targets below are only representative of the following RTSS treatment components: Organ Function, Skills and Habits (only “function-like” and not “activity-like”), and Representations (only those topics found in a systematic review of the voice therapy literature regarding “vocal hygiene”).***

**Vocal Function Targets**

1. Glottal onset (synonym: Glottal attack; Skills & Habits)
  - a. Specify the change on a continuum from breathy to pressed.
  - b. Change in what way: Increase or decrease, improve performance accuracy, increase habitual use of modified voice onset.
  - c. Measurement: Relationship between electroglottogram and microphone signals, clinician auditory judgment.
2. Gross abduction or adduction of the true vocal folds (not during voicing; Organ Functions)
  - a. Change in what way: Increase.
  - b. Measurement: Clinician visual judgment with endoscopy, presence or absence of stridor, patient somatosensory judgment.
3. Loudness (Organ Functions; Skills & Habits)\*
  - a. Types of loudness targets.
    - i. Usual.
    - ii. Range.
    - iii. Variability.
    - iv. Modulation (relevant especially for tremor).
  - b. Change in what way: Increase or decrease, improve performance accuracy, increase habitual use of modified loudness.
  - c. Measurement: Sound pressure level (SPL), shimmer, clinician/patient auditory judgment, rate and extent of SPL modulation.  
*\*The RTSS expects targets to be tied only to a single treatment component group (e.g., only Organ Functions or only Skills & Habits). There appears to be a violation of this expectation as the target “increased loudness” may be directly affected by (1) an Organ Functions mechanism of action (e.g., the nonvolitional Lombard effect) when delivering one ingredient (e.g., applying noise to the ear) or (2) a Skills & Habits mechanism of action (e.g., learning by doing) when delivering a different ingredient (e.g., providing opportunities to practice maximally sustained voicing at maximum loudness). This apparent violation will require future work to determine if this represents two subtly distinct targets with their own mechanisms of action, or a single target affected by two disparate mechanisms of action.*
4. Pitch (Skills & Habits)
  - a. Note: This assumes that the patient is not severely dysphonic; i.e., periodic enough to perceive a pitch.
  - b. Types of pitch targets:
    - i. Usual.
    - ii. Range.
    - iii. Variability.
    - iv. Modulation (relevant especially for tremor).
  - c. Change in what way: Increase or decrease, improve performance accuracy, increase habitual use of modified pitch.
  - d. Measurement: Fundamental frequency ( $f_0$ ), semitones, jitter, # of voice/pitch breaks, clinician/patient auditory judgment, rate and extent of  $f_0$  modulation

5. Supraglottal phonation (Skills & Habits)
  - a. Change in what way: Decrease, increase, improve performance accuracy, decrease or increase habitual use of supraglottal phonation.
  - b. Measurement: Videolaryngoscopy, clinician auditory judgments
6. Vocal fry (Skills & Habits)
  - a. Change in what way: Decrease, improve performance accuracy, decrease habitual use of vocal fry
  - b. Measurement: Creak detector, subharmonics in spectrum/spectrogram/autocorrelation, clinician auditory judgments
7. Voice quality - Breathiness (Skills & Habits)
  - a. Change in what way: Decrease, improve performance accuracy, increase habitual use of modified voice quality
  - b. Measurement: One or more of the following measures: Noise-to-harmonic ratio; cepstral peak prominence; autocorrelation peak; clinician auditory judgments such as Consensus Auditory Perceptual Evaluation: Voice or the Grade-Roughness-Breathiness-Asthenia-Strain scale
8. Voice quality - Roughness (Skills & Habits)
  - a. Change in what way: Decrease, improve performance accuracy, increase habitual use of modified voice quality
  - b. Measurement: One or more of the following measures: Noise-to-harmonic ratio; cepstral peak prominence; autocorrelation peak; clinician auditory judgments such as Consensus Auditory Perceptual Evaluation: Voice or the Grade-Roughness-Breathiness-Asthenia-Strain scale
9. Voice quality - Strain (Skills & Habits)
  - a. Change in what way: Decrease, improve performance accuracy, increase habitual use of modified voice quality
  - b. Measurement: One or more of the following measures: Noise-to-harmonic ratio; cepstral peak prominence; autocorrelation peak; relative fundamental frequency; clinician auditory judgments such as Consensus Auditory Perceptual Evaluation: Voice or the Grade-Roughness-Breathiness-Asthenia-Strain scale.

### Respiratory Function Targets

1. Abdominal movement (Skills & Habits)
  - a. Specify if “during voicing” or “not during voicing.”
  - b. Change in what way: Increase movement, increase smoothness, improve performance accuracy, increase habitual use of modified movement
  - c. Measurement: Magnetometer around abdomen, consistency of abdominal movement rate, plethysmography, inductance bands, ultrasound, patient or clinician tactile (place hand) or visual judgment
2. Clavicular movement (Skills & Habits)
  - a. Specify if “during voicing” or “not during voicing.”
  - b. Change in what way: Decrease movement, improve performance accuracy, increase habitual use of modified movement
  - c. Measurement: Clinician or patient judged visual and/or tactile (place hand) movement of clavicles/shoulders
3. Respiratory coordination for/during vegetative breathing (Skills & Habits)

Definition: Respiratory movements or movement patterns made to modify respiratory efficiency during breathing without voicing or speech.

  - a. Change in what way: Increase smoothness, improve performance accuracy, increase habitual use of modified respiratory motion.
  - b. Measurement: Respiratory kinematics, mean airflow from pneumotachograph (ml/s), # and duration of breath holds, movement of a tissue during exhalation, consistency of expiratory or inspiratory movement rate and duration, clinician- or patient-judged visual and/or tactile (place hand) movement of abdomen/chest/shoulders, plethysmography, inductance bands
4. Respiratory coordination during voicing/speech (Skills & Habits)

Definition: Movements or movement patterns made to modify the interaction between respiratory drive and phonation.

  - a. Change in what way: Improve performance accuracy, increase habitual use of modified respiratory support
  - b. Measurement: Input/output efficiency ratios using sound pressure level divided by subglottal pressure or mean flow during voicing, syllables per exhalation, duration of a breath group, frequency of inhalation during speech, lung volume at phonatory initiation or termination in reference to tidal lung volume, maximum phonation time (MPT), patient or clinician tactile (place hand), visual, or auditory judgment, phonation quotient (vital capacity/MPT)
5. Rib cage movement (Skills & Habits)
  - a. Specify if “during voicing” or “not during voicing.”
  - b. Change in what way: Increase movement, increase smoothness, improve performance accuracy, increase habitual use of modified movement.
  - c. Measurement: Magnetometer around chest, consistency of chest wall movement rate, patient or clinician tactile (place hand) or visual judgment, plethysmography, ultrasound, inductance bands.



### Musculoskeletal Function Targets

1. Alignment (Skills & Habits)
  - a. Specify which muscles/aspects of anatomy.
  - b. Change in what way: Improve, increase.
  - c. Measurement: Clinician judgment during manual palpation, visual observation, objective methods (e.g., stabliograms).
2. Muscle activation levels (Organ Functions; Skills & Habits)\*
  - a. Specify.
    - i. If “during voicing” or “not during voicing.”
    - ii. Which muscles or group of muscles are being targeted? Note, this can include any muscle or muscle group; e.g., anterior neck muscles, expiratory or inspiratory respiratory muscles, muscles around the atlanto-occipital joint, temporomandibular joint, etc.
  - b. Change in what way – Increase or decrease, increase habitual adoption of modified muscle tone.
  - c. Measurement: Clinician manual palpation for activation, patient self-report of muscle activation, symmetry (e.g., thyrohyoid space narrowing or passive hyoid range of motion limitations on the left more than right), passive range of motion (e.g., lateral hyoid range of motion).

*\*There appears to be a violation in the same expectation as above, in that the target “decreased muscle activation levels” may be directly affected by (1) an Organ Functions mechanism of action (e.g., lengthening muscle, changes to the myofascia) when delivering one ingredient (e.g., applying pressure to the muscle) or (2) a Skills & Habits mechanism of action (e.g., learning by doing) when delivering a different ingredient (e.g., providing opportunities to practice modified levels of muscle activation). This apparent violation will require future work as above.*
3. Muscle endurance - Expiratory (Organ Functions)
  - a. Change in what way: Increase.
  - b. Measurement: Sustained maximum voluntary ventilation, incremental threshold loading (breathing through a device that has some resistance against respiration and the resistance is gradually increased over time until the subject cannot successfully breathe against the resistance. This failure pressure is thought to be a measure of respiratory endurance).
4. Muscle endurance - Inspiratory (Organ Functions)
  - a. Change in what way: Increase.
  - b. Measurement: Sustained maximum voluntary ventilation, incremental threshold loading (breathing through a device that has some resistance against respiration and the resistance is gradually increased over time until the subject cannot successfully breathe against the resistance. This failure pressure is thought to be a measure of respiratory endurance).
5. Muscle strength - Expiratory (Organ Functions)
  - a. Change in what way: Increase.
  - b. Measurement: Maximum expiratory pressure, sustained maximum voluntary ventilation, breathing through a device that has some resistance against expiration
6. Muscle strength - Inspiratory (Organ Functions)
  - a. Change in what way: Increase.
  - b. Measurement: Maximum inspiratory pressure, sustained maximum voluntary ventilation, breathing through a device that has some resistance against inspiration
7. Range of motion - Passive (Organ Functions)
  - a. Specify which joints or anatomical area.
  - b. Change in what way: Increase.
  - c. Measurement – Clinician manual palpation, visual observation, measurement of displacement (mm, cm, inches)
8. Range of motion - Active (Organ Functions)
  - a. Specify which joints or anatomical area.
  - b. Change in what way: Increase.
  - c. Measurement: Clinician manual palpation, visual observation, measurement of displacement (mm, cm, inches)
9. Vocal endurance (Organ Functions)
  - a. Change in what way: Increase.
  - b. Measurement: Patient self-report of vocal status changes in daily life or voice dosimeter/monitoring to estimate amount of voicing before vocal fatigue.

### Somatosensory Function Targets

1. Resonance (Skills & Habits)
  - a. Description of the desired resonance.
    - i. Anatomy: Oral, nasal, facial, chest, throat.
    - ii. Direction: Forward, backward, higher/lighter/lift, lower.
    - iii. Timbre: Bright, dark, twang.
  - b. Change in what way.
    - i. Increase or decrease: Amount of requested resonance.
    - ii. Increase or decrease focus: Amount to which requested resonance adheres to only the description requested.
    - iii. Improve performance accuracy, increase habitual use of modified resonance.
  - c. Measurement: Clinician auditory judgment, ratio or slope of spectral energy, amplitude of high frequency energy in spectrum
2. Kinesthetic discrimination (Skills & Habits)
  - a. Anatomical location of desired vibrotactile sensation: Mask, mouth, nose, face, alveolar ridge, back of throat, chest, etc.
  - b. Change in what way.
    - i. Increase or decrease: Amount of vibrotactile sensation in anatomical location requested.
    - ii. Increase or decrease focus: Amount to which requested vibrotactile sensation adheres to only the anatomical location requested.
    - iii. Improve performance accuracy, improve judgment accuracy.
  - c. Measurement: Patient self-report of where vibrations are occurring, clinician judgment via touch to anatomical location (if possible), amplitude of accelerometer waveform on specific anatomical location, Eulerian Video Magnification of face, neck, or chest.
3. Pain/discomfort/soreness (Organ Functions)
  - a. Specify the location of pain or discomfort.
  - b. Change in what way: Decrease.
  - c. Measurement: Patient self-report.

### Auditory Function Targets

1. Voice quality discrimination (Skills & Habits)
  - a. Specify.
    - i. What is being judged: Self-monitoring vocal productions, non-self models from clinician, recordings, etc.
    - ii. Type of voice quality: Modal, overall dysphonia, strain, breathiness, roughness, vocal fry.
  - b. Change in what way: Improve judgment accuracy.
  - c. Measurement: Patient self-report compared with clinician auditory judgment, a combination or single use of the objective measures like cepstral peak prominence, spectral tilt, H1-H2, open quotient, etc.
2. Pitch discrimination (Skills & Habits)
  - a. Note: Assumes that the patient is not severely dysphonic; i.e., periodic enough to perceive a pitch.
  - b. Specify.
    - i. What is being judged: Self-monitoring vocal productions, non-self models from clinician, recordings, etc.
    - ii. Type of pitch monitoring target: Habitual, range, variability, modulation.
  - c. Change in what way: Improve judgment accuracy.
  - d. Measurement: Patient self-report compared with clinician auditory judgment or objective measures (e.g., f0, jitter)
3. Loudness discrimination (Skills & Habits)
  - a. Specify.
    - i. Self-monitoring vocal productions, non-self models from clinician, recordings, etc.
    - ii. Type of loudness monitoring target: Habitual, range, variability, modulation.
  - b. Change in what way: Improve judgment accuracy.
  - c. Measurement: Patient self-report compared with clinician auditory judgment or objective measures (e.g., SPL, shimmer).

**Pedagogy and Counseling Targets**

1. Voice and vegetative laryngeal use strategies; examples include voice rest, modified voice rest, decreased loudness in specific situations, coughing or throat clearing, etc.
2. Reflux strategies
3. Hydration strategies
4. Recreational drug use; examples include modifications to using alcohol, caffeine, smoking, vaping, etc.

For all 4 targets above:

- a. Pedagogical target (Representations).
  - i. Change in what way: Increased amount of knowledge, Improved accuracy of knowledge.
  - ii. Measurement: Patient verbal recall of information, written quiz, verbal questioning, demonstration of “how to” knowledge.
- b. Counseling target (Representations).
  - i. Change in what way: Modified beliefs and/or values, enhanced motivation, modified attitudes (increased positive or negative attitude toward...).
  - ii. Measurement: Patient report of their attitudes, motivation, and beliefs; ambulatory monitoring of psychological state (self-reported affect or emotion throughout a day) or psychophysiological state (heart rate variability, electrodermal skin activity).
- c. Habit formation target (Skills & Habits).
  - i. Change in what way: Decreased effort to implement, formation of habit.
  - ii. Measurement: Patient report of cognitive effort/automaticity, ambulatory monitoring of vocal function in specific circumstances.

**Speech & Communication Targets**

The patient functions outlined here will often be Aims of voice therapy, but there are instances where these functions would be targets. For example:

- A patient where large improvements in phonation are not likely to be elicited through behavioral approaches alone (e.g., vocal fold scar, spasmodic dysphonia, essential tremor, bilateral vocal fold paralysis) may be given ingredients such as “speak face-to-face” or “slow down/speed up speech” or “exaggerate speech” to increase intelligibility or comprehensibility.
    1. Intelligibility (Skills & Habits).
      - a. Change in what way: Improved performance accuracy, increased, formation of habit.
      - b. Measurement: Clinician judgment (e.g., perceptual measurement of intelligibility); vowel space area; formant tracking metrics.
    2. Comprehensibility (Skills & Habits).
      - a. Change in what way: Improved performance accuracy, increased, formation of habit.
      - b. Measurement: Clinician judgment
-

---

## Appendix B (p. 1 of 7)

### Final List of RTSS-Voice Ingredients

---

Generally, note that there are no “mechanisms of action” in ingredient labels

---

Since “mechanisms of action” (how or why the ingredients cause changes in the target) are hypothesized (and frequently have limited empirical evidence to support them), any reference to a mechanism of action in an ingredient label is purposely minimized. Examples of “mechanisms of action” being used as ingredient labels include: sensory-motor perturbation, confusion, myofascial release, distracting the patient, reposturing the larynx, impedance-matching, creation of a phonatory configuration with less potential for trauma, and thyroarytenoid–cricothyroid balance/engagement.

**Also, note that delivery vehicles for ingredients may be chosen based on the following criteria (using SOVT’s as an example):**

- Treatment theory (e.g., straws provide more explicit feedback about airflow than humming).
- Volition such as patient preference, capacity, stimulability, etc. (e.g., the patient is more likely to practice outside the clinic with an /m/ than a straw because there is no need to carry a straw with them).
- Clinician preference or comfort with a particular SOVT is not a reason to theoretically distinguish between various types of SOVTs.

### Nonvolitional ingredients

**Generally, note that nonvolitional ingredients can be provided alone to passively change patient functioning or provided with additional ingredients to actively change patient functioning.** According to the RTSS, ingredients such as devices, providing energy to tissue, or applying pressure (etc.) can be used to passively or actively affect a target:

- i. *If the ingredient passively achieves an associated target, it is considered to have a treatment target of its own (no additional ingredients necessary).* For example:
  - Provide a voice amplifier (ingredient) to increase loudness (target) in a patient with bilateral vocal fold paralysis.
  - Apply kneading pressure in the thyrohyoid space (ingredient) to decrease muscle activation levels at rest as measured by increased hyoid passive range of motion (target) in a patient with a unilateral vocal fold polyp.
- ii. *As is more often the case, a patient must develop the skill of using a device appropriately or using their voice appropriately with the clinician-applied ingredient. Therefore, the ingredient DOES NOT have a target of its own and must be combined with some opportunity to practice.* For example:
  - Provide a voice amplifier (ingredient 1) and opportunities to practice voicing in a certain way with the amplifier (ingredient 2) to improve vocal endurance (target) in a teacher with vocal fold nodules.
  - Apply downward pressure on the hyoid bone (ingredient 1) during opportunities to practice voicing in a certain way (ingredient 2) to decrease strained voice quality (target) in a patient with primary muscle tension dysphonia

1. **Apply heat:** Apply heat to (i) an anatomical location (ii) with a delivery vehicle and (iii) at a specific dose.

*Delivered in what way (specify only those that are theoretically relevant):*

- i. Anatomical location: List structures that were targeted by the heat.
- ii. Delivery vehicle: Delivery vehicles can include towel, rubber water bottle, compress, etc.
- iii. Dose: Dose can be measured by the degree of heat (e.g., warm to the touch, a specific temperature), time applied (per repetition or total time), repetitions, and/or timing of repetitions (e.g., cooling time between applications).

*Possible targets can include one or more of the following (this listing is not exhaustive):* Decreased passive muscle tone; decreased pain; increased passive range of motion; etc.

2. **Apply low level light:** Apply (i) a wavelength of light on (ii) an anatomical location at (iii) a specific dose.

*Delivered in what way (specify only those that are theoretically relevant):*

- i. Wavelength: Typically, either (or both) of two light wavelengths are used: infrared and/or red.
- ii. Anatomical location: List structures that were targeted by the light.
- iii. Dose: Dose will include descriptions such as the duration of time and the amount/intensity of light that was applied, as well as the schedule of repetition (e.g., the number of repetitions, how the repetitions were structured in time).

*Possible targets can include one or more of the following (this listing is not exhaustive):* Decreased passive muscle tone; decreased pain/discomfort; etc.

*Note:* When infrared light is used, it is possible to deliver two ingredients simultaneously (light and heat).



3. **Apply noise:** Apply (i) a type of noise, (ii) with a delivery vehicle, (iii) at a certain dose during voicing.  
*Delivered in what way (specify only those that are theoretically relevant):*
  - i. Type of noise: These include categories of noise (e.g., cocktail, white, pink, brown), environmental sounds (e.g., music, air conditioners, reverb), etc.
  - ii. Delivery vehicle: Delivery vehicles include headphones, speakers, natural environments, acoustics of natural or virtual environments, etc.
  - iii. Dose: This is the amount of noise provided such as one or a combination of the following (depending upon what was theoretically relevant to the treatment): intensity/loudness (e.g., dB SPL, dB-A, dB-C, sones), at various frequencies (depends upon the type of noise), and/or the ratio of intensity versus the patient's vocal intensity (e.g., signal-to-noise ratio).*Possible targets can include one or more of the following (this listing is not exhaustive):*  
Nonvolitional changes in voicing such as increased loudness (e.g., a patient with Parkinson's disease); decreased strained voice quality (e.g., a patient who is functionally aphonic), etc.
4. **Apply physical occlusion to ear(s):** The clinician or patient (i) use a delivery vehicle to (ii) provide a dose or degree of sound attenuation to the external auditory canal.  
*Delivered in what way (specify only those that are theoretically relevant):*
  - i. Delivery vehicle: Delivery vehicles include fingers, ear plugs, etc.
  - ii. Dose: Amount of sound attenuation can be described by degree (partial, complete), unilateral versus bilateral, and/or the sound attenuation properties of the materials occluding the external auditory canal (e.g., ear plugs with higher or lower noise reduction ratings).*Possible targets can include one or more of the following (this listing is not exhaustive):* Increased loudness; if this is combined with vocal practice, targets could be increased vocal endurance; decreased strained voice quality; etc.
5. **Apply pressure:** Apply force with (i) a delivery vehicle, (ii) in a certain manner, (iii) on an anatomical location, (iv) during a context, (v) at a certain dose.  
*Delivered in what way (specify only those that are theoretically relevant):*
  - i. Delivery vehicle: This could be specified as manual (e.g., specific finger/thumb combination, palm of hand) or an external device (e.g., vibrator).
  - ii. Manner: Check all that apply to the treatment provided and describe them: kneading (circular), stroking (uni- or bidirectional), static, pulling in one direction (specify the direction like lateral, inferior, etc.), and oscillation (e.g., gentle shaking around a set point, repetitive pushing and releasing anterior pressure, alternatively pulling left and then right).
  - iii. Anatomical location: List structures that were targeted by the pressure.
  - iv. Context: This could be specified as during rest, voicing, breathing, a specific bodily orientation (lying down, sitting up), etc.
  - v. Dose: Dose could include descriptions such as the amount of pressure applied (e.g., visual indicators like blanching of the fingertips or depth like superficial/deep), number of repetitions, the amount of time pressure was applied, the timing at which pressure was applied, and/or some measure of total force delivered. For example, the duration of pressure often depends upon (1) when the excess muscle activation is minimized or stops or (2) if combined with opportunities to practice something like voicing or breathing, it will depend upon when voicing occurs, the duration of voicing, when voicing or breathing improves, etc.*Possible targets can include one or more of the following (this listing is not exhaustive):* Nonvolitional changes in patient functioning like decreased passive muscle tone; increased passive range of motion; decreased strained voice quality; decreased pitch; if combined with opportunities to practice, targets could include decreased strained, breathy, or rough voice quality; decreased pitch; etc.
6. **Apply topical numbing:** Provide (i) a numbing agent (ii) using a delivery method to introduce the agent (iii) to an anatomical location (iv) at a specific dose.  
*Delivered in what way (specify only those that are theoretically relevant):*
  - i. Numbing agent: If a specific numbing substance was chosen because of treatment theory or patient volition, this should be specified (e.g., cetacaine, lidocaine).
  - ii. Delivery method: Describe the delivery method (e.g., spray, injection, drip, air pressure).
  - iii. Anatomical location: List the structures that were targeted by the anesthesia.
  - iv. Dose: Dose could include the intensity of the numbing substance and the amount delivered (e.g., 1% lidocaine in 20 ml). It may also include the number of times the substance was provided, the time duration of administration, the amount of air pressure used (if using air pressure to deliver substance). Dose provided may also be dependent upon patient self-reported level of numbness (e.g., complete, partial, minimal).*Possible targets can include one or more of the following (this listing is not exhaustive):* Nonvolitional changes in patient functioning like decreased pain; decreased sensation; nonvolitionally improved voice quality; decreased supraglottal phonation; decreased chronic cough; decreased throat clearing; etc.

7. **Provide semi-occluded vocal tract (SOVT) postures:** Ask the patient to voice with a (i) delivery vehicle that narrows their vocal tract (ii) and creates a dose of resistance.

*Delivered in what way (specify only those that are theoretically relevant):*

- i. Delivery vehicle: The clinician should specify whether the SOVT delivery vehicle was:
  - a. An external device: The device (e.g., straw, mask, kazoo, wax paper, flowball device) and the treatment-relevant properties of the device should be described (e.g., was the straw submerged in water, length/width of straw, size of mask).
  - b. Anatomical: Lip trill, tongue trill, raspberry, inverted megaphone, semi-occluded vowels such as /u, o, i/, voiced fricatives such as /z, v/, nasal consonants such as /m, n/, etc.
- ii. Dose: This would include the amount of resistance associated with the SOVT used; and could be indirectly specified as the amount of occlusion in the vocal tract (diameter of semi-occlusion, or complete labial closure, or +/- straw submerged in water).

*Possible targets can include one or more of the following (this listing is not exhaustive):* Nonvolitional changes in voicing such as decreased strained voice quality and improved respiratory coordination during voicing, etc.

8. **Provide voice amplification:** Provide the patient with (i) a device with clinically meaningful characteristics to (ii) use during prescribed situations (iii) where a dose of amplification is necessary.

*Delivered in what way (specify only those that are theoretically relevant):*

- i. Device characteristics: Specify characteristics of the device that are considered relevant to achieving the target (e.g., portable device; dedicated wireless device such as frequency modulation or Bluetooth; microphone characteristics such as quality, ease of wear, location of wear).
- ii. Prescribed situations for use: This could include descriptions such as: all the time, during work, during leisure, etc.
- iii. Dose: This is the amount of amplification provided, which could be measured in unweighted or weighted decibels (dB); e.g., “sound pressure level (SPL)” or A-weighting (dB-A), C-weighting (dB-C), phons, sones, etc. This dose may also be represented as some ratio of “signal” (amount of amplification) versus “noise” (amount of background noise).

*Possible targets can include one or more of the following (this listing is not exhaustive):* Nonvolitional changes in voicing such as increased loudness (e.g., for a patient with bilateral vocal fold paralysis); if combined with opportunities to practice targets could include decreased loudness (e.g., for a patient with bilateral vocal fold nodules), etc.

### **Ingredients for Direct Targets That Involve Patient Volition**

9. **Gross vocal fold adduction exercises (without voicing):** Opportunity to perform a movement (i) associated with maximal vocal fold adduction and (ii) a delivery method to (iii) generate a dose of increased pressure in the thoracic cavity/lower airway, as well as (iv) a rule(s) for progression of the subglottal resistance.

*Delivered in what way (specify only those that are theoretically relevant):*

- i. Movement associated with maximal vocal fold adduction: Check all that are theoretically relevant to the treatment provided and describe them: hard glottal onset, breath holding at the level of the true vocal folds, a swallow maneuver (e.g., supraglottic swallow), glottal clicks.
- ii. Delivery method: Check all that are theoretically relevant to the treatment provided and describe them: pushing, pulling, leaning against an object, lifting something with a specified weight.
- iii. Dose: Dose includes both the number of repetitions and sets, and some amount of resistance. The level of resistance can be based on what the patient maximally tolerates (i.e., failure or intolerance would be measured by air leaking into the upper airway) or can be indirectly measured via weight that must be picked up while holding Valsalva, or amount of displacement on the abdomen or weight on the abdomen during Valsalva.
- iv. Progression Rule(s): As the patient improves, the challenge level will be increased in a specific way. For example, when the patient can perform the prescribed number of repetitions and sets of repetitions at the desired resistance, how will the resistance be adjusted?

*Possible targets can include one or more of the following (this listing is not exhaustive):* Increased gross adduction of the true vocal folds.

10. **Provide feedback:** Clinician provides information on patient performance through decisions regarding (i) who delivers the feedback, (ii) type of feedback, (iii) timing of feedback, (iv) feedback modality (or multiple modalities), and (v) dose.

*Delivered in what way (specify only those that are theoretically relevant):*

- i. Who delivers the feedback: Check all that are theoretically relevant and describe: the clinician, a third party (caregiver, teacher, friend, etc.), patient self-evaluation or the feedback could be a comparison between the patient's and the clinician's perception of accuracy/error.
- ii. Type of feedback: Check all that are theoretically relevant and describe: modeling; a continuum between knowledge of results versus knowledge of performance; different types of scales such as binary (good vs. bad); categorical (better, OK, worse); interval (0–100 scale where 0 is bad and 100 is good); ratio (current performance in relation to a past reference of performance); progress tracking (log of practice/exercise or performance during practice/exercise); augmented feedback (e.g., biofeedback).

## Appendix B (p. 4 of 7)

### Final List of RTSS-Voice Ingredients

- iii. Timing of feedback: Is the feedback delivered during performance (knowledge of performance) or after performance (knowledge of results)? When feedback is after performance, if relevant to the treatment, specify the amount of time after performance (seconds, minutes, hours, etc.).
  - iv. Feedback modality: Check all that are theoretically relevant and describe: e.g., visual, verbal, physical/tactile, etc.
  - v. Dose: If applicable, specifications would quantify frequency (e.g., after every trial or only after every 5<sup>th</sup> trial); based on performance (e.g., until attain 90% accuracy); categorical (no feedback, moderate feedback, maximum feedback), amount of feedback related to optimal challenge (e.g., fading, amount of feedback based on maintaining 80% accuracy).  
*Possible targets can include one or more of the following (this listing is not exhaustive): All targets involving volitional patient behavior can have feedback ingredients.*
11. **Provide opportunities to practice alignment/posture:** Provide the patient with opportunities to practice (i) a specific template of alignment/posture (ii) on a continuum of variability/difficulty (iii) for a prescribed dose, and (iv) progression.  
*Delivered in what way (specify only those that are theoretically relevant):*
- i. Template of alignment/posture: Check all that are theoretically relevant and describe them: position (sitting, standing, lying, etc.), anatomical locations to modify (knees, feet, shoulders, chin, stomach, etc.).
  - ii. Variability/difficulty.
    - Practice variability: Describe how practice was structured such as blocked, alternating, variable, negative (alternate between voicing in a desired manner and the patient’s baseline manner).
    - What was practiced in a variable way (and how much variability): Describe what was intentionally varied by a specific amount for treatment purposes, such as generalization (e.g., variable body positions, varying circumstances such as singing or different types of speech).
  - iii. Dose: Dose includes the number of opportunities to practice, total number of practice repetitions, and/or the practice schedule (e.g., massed vs. spaced, blocked vs. variable).
  - iv. Progression Rule(s): As the patient’s skill improves, the challenge level will be increased in a specific way (e.g., practice at a difficulty level until the patient attains a performance criterion such as “80% accuracy”). Please describe how difficulty is to be increased: more difficult body positions, cognitive load during alignment/postural practice (e.g., topics requiring more or less cognitive effort), affective load during alignment/postural practice (e.g., situations or topics with more or less stress, emotional connection, etc.).  
*Possible targets can include one or more of the following (this listing is not exhaustive): Improved active alignment; changes in any voicing-related target (if combined with opportunities to practice voicing); changes in any respiration-related target (if combined with opportunities to practice breathing); etc.*
12. **Provide opportunities to practice breathing:** Provide the patient with opportunities to practice (i) a specific template of breathing (ii) on a continuum of variability/difficulty (iii) for a prescribed dose and (iv) progression.  
*Delivered in what way (specify only those that are theoretically relevant):*
- i. Template of breathing: Check all that are theoretically relevant and describe them: during vegetative breathing/voicing/speech, oral versus nasal breathing, airflow without voicing, kinematics such as clavicular/rib cage/abdominal movement, duration, timing in relation to some reference such as resting expiratory volume or voicing initiation, constriction of the upper airway used during breathing (should describe the constriction; e.g., pursed lips, hand in front of mouth, straw), breath holding, rate, body position (supine, sitting, standing, etc.).
  - ii. Variability/difficulty.
    - Practice variability: Describe how practice was structured such as blocked, alternating, variable, negative (alternate between breathing in a desired manner and the patient’s baseline manner), etc.
    - What was practiced in a variable way (and how much variability): Describe what was intentionally varied by a specific amount for treatment purposes, such as generalization (e.g., variations in abdominal movement, rate of breathing, or duration of breathing).
  - iii. Dose: Dose includes the number of opportunities to practice, total number of practice repetitions, and/or the practice schedule (e.g., massed vs. spaced, blocked vs. variable).
  - iv. Progression Rule(s): As the patient’s skill improves, the challenge level will be increased in a specific way (e.g., practice at a difficulty level until the patient attains a performance criterion such as “80% accuracy”). Describe how difficulty is to be increased: speech/nonspeech complexity during breathing practice (e.g., vowels vs. spontaneous speech), what about breathing is more difficult (faster or slower rate, environment (e.g., presence of noxious agent associated with more difficult breathing), cognitive load during breathing practice (e.g., topics requiring more or less cognitive effort), affective load during breathing practice (e.g., situations or topics with more or less stress, emotional connection, etc.), physical exertion during breathing practice (e.g., walking, jogging, sprinting, dribbling a basketball).  
*Possible targets can include one or more of the following (this listing is not exhaustive): Improved respiratory coordination during voicing/during vegetative breathing; improved abdominal or rib cage or clavicular movement during breathing; etc.*

13. **Provide opportunities to practice modified levels of muscle activation:** Provide the patient with opportunities to practice (i) a specific template of increased and/or decreased muscle activation (ii) in a muscle(s)/muscle group(s) (iii) on a continuum of variability/difficulty, (iv) in a specific context, (v) for a prescribed dose, and (vi) progression.  
*Delivered in what way (specify only those that are theoretically relevant):*
- i. Template of increased and/or decreased muscle activation: Check all that are theoretically relevant and describe them: amplitude of muscle activation requested (high and low), duration of time in high contraction and relaxed state.
  - ii. Muscle/muscle group: List muscle/muscle groups that were volitionally increased a decreased in activation.
  - iii. Variability/difficulty.
    - Practice variability: Describe how practice was structured, such as blocked, alternating, variable, etc.
    - What was practiced in a variable way (and how much variability): Describe what was intentionally varied by a specific amount for treatment purposes, such as generalization. Examples include variation in muscles targeted (e.g., anterior neck, then posterior neck, then jaw) or amount/timing of maximum activation/maximum de-activation cycle (e.g., more or less activation or hold the posture for a longer or shorter time durations).
  - iv. Context: This can be specified as during rest, voicing, breathing, a specific bodily orientation (lying down, sitting up), etc.
  - v. Dose: Dose includes the number of opportunities to practice, total number of practice repetitions, and/or the practice schedule (e.g., massed vs. spaced, blocked vs. variable).
  - vi. Progression Rule(s): As the patient’s skill improves, the challenge level will be increased in a specific way (e.g., practice at a difficulty level until the patient attains a performance criterion such as “80% accuracy”). Please describe how difficulty is to be increased: voicing or not voicing during practice, different body positions (reclined vs. standing), cognitive load during practice (e.g., topics that are more or less cognitively demanding), affective load during practice (e.g., situations or topics with more or less stress, emotional connection, etc.).
- Possible targets can include one or more of the following (this listing is not exhaustive):* Decreased muscle activation in (specify muscle/muscle group).
14. **Provide opportunities to practice sensory discrimination:** Provide the patient with opportunities to practice (i) a specific template of sensory discrimination (ii) with a signal from the patient or from another source, (iii) on a continuum of variability/difficulty (iii) for a prescribed dose, and (iv) progression.  
*Delivered in what way (specify only those that are theoretically relevant):*
- i. Template of sensory discrimination: Check all that are theoretically relevant and describe them: pitch, loudness, voice quality (choose roughness, breathiness, strain), registration (choose fry, head voice, chest voice), vibrotactile sensation (at which anatomical location(s)), airflow, pressure, respiration, +/- semi-occluded vocal tract, visual identification of movement (e.g., jaw opening, clavicular movement), discriminating between higher versus lower levels of muscle tone, etc.
  - ii. Source of signal: Check all that apply and describe them: patient generated (phonation, breathing, resting, etc.), clinician generated (phonation, breathing, resting, etc.), external devices (speaker, headphones, vibrator, etc.)
  - iii. Variability/difficulty.
    - Practice variability: Describe how practice was structured such as blocked, alternating, variable, negative (alternate between a desired behavior and the patient’s baseline behavior).
    - What was practiced in a variable way (and how much variability): Describe what was intentionally varied by a specific amount for treatment purposes, such as generalization (e.g., pitch variation, loudness variation, voice quality variation).
  - iv. Dose: Dose includes the number of opportunities to practice, total number of practice repetitions, and/or the practice schedule (e.g., massed vs. spaced, blocked vs. variable).
  - v. Progression Rule(s): As the patient’s skill improves, the challenge level will be increased in a specific way (e.g., practice at a difficulty level until the patient attains a performance criterion such as “80% accuracy”). Please describe how difficulty is to be increased: addition of noise to make the sensory signal harder to detect (e.g., environmental noise), cognitive load during sensory discrimination practice (e.g., topics requiring more or less cognitive difficulty), affective load during sensory discrimination practice (e.g., situations or topics with more or less stress, emotional connection, etc.).
- Possible targets can include one or more of the following (this listing is not exhaustive):* Any auditory function target, improved kinesthesia, etc.
15. **Provide opportunities to practice voicing:** Provide the patient with opportunities to practice (i) a specific template of voicing (ii) on a continuum of variability/difficulty (iii) for a prescribed dose and (iv) progression.  
*Delivered in what way (specify only those that are theoretically relevant):*
- i. Template of voicing: Check all that are theoretically relevant and describe them: loudness, pitch, sustained phonation, airflow (i.e., “flow phonation”), subglottal pressure, periodicity, inhalation phonation, supraglottal phonation, registration (choose fry, head voice, chest voice), glottal onset, vegetative vocalizations, resonance (e.g., “forward resonance” or “twang”), half-swallow boom, +/- semi-occluded vocal tract (if so, select and specify the SOVT ingredient), speech material (prolonged vowels, nonspeech vowel–consonant combinations, spontaneous speech, etc.), rate of production (fast, slow, patterns of fast/slow, etc.).



## Appendix B (p. 6 of 7)

### Final List of RTSS-Voice Ingredients

- ii. Variability/difficulty.
  - Practice schedule: Describe how practice was structured such as blocked, alternating, variable, negative (i.e., alternate between voicing in a desired manner and the patient's baseline manner), etc.
  - What was practiced in a variable way (and how much variability): Describe what was intentionally varied by a specific amount for treatment purposes, such as generalization (e.g., variation in pitch, loudness).
- iii. Dose: Dose includes the number of opportunities to practice, total number of practice repetitions, and/or the practice schedule (e.g., massed vs. spaced, blocked vs. variable).
- iv. Progression Rule(s): As the patient's skill improves, the challenge level will be increased in a specific way (e.g., practice at a difficulty level until the patient attains a performance criterion such as "80% accuracy over # trials"). Describe how difficulty is to be increased: speech/nonspeech complexity (e.g., vowels vs. spontaneous speech), what about the voicing is more difficult (e.g., softer-than-comfortable), environment (e.g., environmental noise levels, room acoustics), cognitive load (e.g., topics requiring more or less cognitive difficulty), affective load (e.g., situations or topics with more or less stress, emotional connection, etc.).

*Possible targets can include one or more of the following (this listing is not exhaustive):* Increased/decreased habitual pitch or loudness; decreased strained voice quality; increased forward resonance; etc. (i.e., all targets requiring improvements in voicing).

16. **Resistance exercises - Inspiratory:** Opportunity to (i) perform an inspiratory movement (ii) with a device that uses a specific method to provide a (iii) dose of resistance against inhalation as well as (iv) a rule(s) for progression of the resistance level.

*Delivered in what way (specify only those that are theoretically relevant):*

- i. Inspiratory movement: Check all that are theoretically relevant and describe them: kinematics such as clavicular/rib cage/abdominal movement, duration, body position (prone, supine, lying on side, standing, sitting, etc.).
- ii. Method of applying resistance: Specific methods include passive flow-resistance devices (resistance level based on the diameter of tube being used), pressure threshold devices (resistance based on a valve that blocks airflow unless a pressure threshold is exceeded), etc.
- iii. Dose: Dose includes both the number of repetitions, the number and schedule of sets prescribed, and some amount of resistance. The level of resistance is typically normalized according to some patient-specific reference value like the percentage of the patient's maximum inspiratory resistance.
- iv. Progression Rule(s): As the patient improves, the challenge level will be increased in a specific way (e.g., when patient can perform the prescribed number of repetitions and sets of repetitions at the desired resistance level, how will the resistance level be adjusted?).

*Possible targets can include one or more of the following (this listing is not exhaustive):* Increased inspiratory muscle strength or endurance.

17. **Resistance exercises - Expiratory:** Opportunity to (i) perform an expiratory movement, (ii) with a device that uses a specific method provide a (iii) dose of resistance against exhalation, as well as (iv) a rule(s) for progression of the resistance level.

*Delivered in what way (specify only those that are theoretically relevant):*

- i. Expiratory movement: Check all that are theoretically relevant and describe them: kinematics such as clavicular/ribcage/abdominal movement, duration, body position (prone, supine, lying on side, standing, sitting, etc.).
- ii. Method of applying resistance: Specific methods include passive flow-resistance devices (resistance level based on the diameter of tube being used), pressure threshold devices (resistance based on a valve that blocks airflow unless a pressure threshold is exceeded), etc.
- iii. Dose: Dose includes both the number of repetitions, the number and schedule of sets prescribed, and some amount of resistance. The level of resistance is typically normalized according to some patient-specific reference value like the percentage of the patient's maximum expiratory resistance.
- iv. Progression Rule(s): As the patient improves, the challenge level will be increased in a specific way (e.g., when patient can perform the prescribed number of repetitions and sets of repetitions at the desired resistance level, how will the resistance level be adjusted?).

*Possible targets can include one or more of the following (this listing is not exhaustive):* Increased expiratory muscle strength or endurance.

**Representations ingredients:**

18. **Provide vocal hygiene information:** Clinician provides (i) bits of information (ii) through a modality (or multiple modalities) and (iii) a method (or multiple methods), (iv) at a dose.

*Delivered in what way (specify only those that are theoretically relevant):*

- i. Information bits: Check any that are theoretically relevant and describe them:
  - a. Effects of actions/activities/substances related to voice difficulties like coughing/throat clearing, extreme laughing or crying, excessive voice use, hard glottal onset/attacks, hydration, producing unconventional sounds, reflux, sleep, talking/singing over noise, use of recreational drugs, medications that interfere with vocal function, etc.
  - b. Patient's diagnosis(s) and related anatomical/physiological changes.
  - c. Treatment and prognosis/expectations for treatment.
- ii. Modality: Check all that are theoretically relevant such as visual, verbal, written, etc.
- iii. Method: Check all that are theoretically relevant such as didactic, discussion, Socratic, etc.
- iv. Dose: If applicable, specifications would quantify the number of informational bits provided, # of bits per second, minute, hour session, and bits may include the # of bullet points covered, depth of information on a specific topic, repetitions or amount of rehearsal/review of information, etc. The difficulty of the informational bits may also be included.

*Possible targets can include one or more of the following (this listing is not exhaustive):* All Representation Targets can have these ingredients: e.g., increased amount of knowledge regarding reflux strategies, enhanced motivation to use hydration strategies in daily life, modified beliefs about how smoking effects the voice.

19. **Provide volition ingredient(s):** The clinician provides (i) bits of information to enhance the patient's Capability, Opportunity, and Motivation to perform a desired Behavior (COM-B) at (ii) a certain dose.

*Delivered in what way (specify only those that are theoretically relevant):*

- i. Information bits: COM-B ingredients could include:
  - a. Changing knowledge through: providing a template of what is to be practiced or what behavior is desired; provision of didactic information (various modalities such as written, verbal, visual, etc.); prompting patient to acquire information, information-organizing methods (e.g., chunking); repetition/prompting rehearsal of information; Socratic methods (question-answer format); mnemonic aids, modeling, cueing, etc.
  - b. Changing attitude (propensity to act) through: Provision of appeals based on values, norms, fear, etc.; reassurance; promotion of alternative interpretations; elicitation of change talk (i.e., Motivational Interviewing), etc.
  - c. Changing motivation/effort through: Provision of rationale(s) (e.g., for treatment or treatment activity); persuasion, bargaining, contracting; methods to instill trust in clinician (rapport, credibility); use of patient's preferred tasks or materials; goal setting with or for patient; reinforcement (positive, negative), incentives, punishment, etc.
  - d. Enhancing the patient's opportunity to perform the desired behavior by: Prompting problem-solving to ensure adequate space/support/other resources to support performance of volitional activity, collaborative scheduling of volitional activity, etc.
- ii. Dose: If applicable, specifications would quantify the number of informational bits provided, # of bits per second, minute, hour session, and bits may include the # of bullet points covered, depth of information on a specific topic, repetitions or amount of rehearsal/review of information, etc. The difficulty of the informational bits may also be included.

*Possible targets can include one or more of the following (this listing is not exhaustive):* If the volitional treatment is supervised, then the clinician will list these ingredients without a separate volition target. If the volitional treatment is unsupervised, then the ingredients will always be listed with a volition target phrased "performance of [insert treatment activity] as directed."

---