This article examines the nature, uses, and limitations of the large variety of existing, so-called, hypnosis scales; that is, instruments that have been proposed for the assessment of hypnotic behavior. Although the major aim of most of the scales ostensibly seems to be to assess several aspects of hypnotic states, they are found generally to say little about these and much more about responses to suggestions. The greatest application of these scales is to be found in research, but they also have a limited place in clinical work.

As of 1998, at the time I was working on the second edition of my Practice of Hypnotism (Weitzenhoffer, 2000) and preparing the chapter on hypnotic depth, its measurement and related subjects, there existed at least 25 instruments variously called scales of “hypnotic depth,” “of hypnotic susceptibility,” “of suggestibility,” “of hypnotizability,” and variations thereof. I am fairly sure I missed other existing ones and, most likely, new ones have cropped up since then. A listing, description and a detailed discussion of most of these instruments will be found in the above mentioned work and will not be repeated here.

Why so many? Do we need all of them? If so, how does one go about deciding which ones to use and when to use them?

A Note On Terminology

The language of hypnotism contains a great deal of ambiguity that can easily lead to misunderstandings regarding that which is being communicated. Typically the term “hypnosis” has long and widely been used to denote a special condition of being, but, increasingly during the last fifty years or so, it has also come to denote something quite different; namely, the production, study and use of so-called hypnotic phenomena, not a few of which, in my opinion, doubtfully are so. But then here too there exists an unfortunate ambiguity. This is really not a new problem, for the development of a proper nomenclature for hypnotism was one of the key items on the agenda of the First International Congress of Hypnotism that took place in Paris in 1889. Nothing came out of this. Considering this situation it may be of value to clarify from the outset how some key terms will be used in this article. Except then
for situations that obviously require some other, conventional, usage, such as when I refer to the “Society for Clinical Hypnosis,” or to an article titled “Theories of Hypnosis,” I will otherwise generally reserve the term hypnosis for the state and the term hypnotism for the production, study, and use of suggestion with the state of hypnosis presumably being present, whether or not it adds anything tangible to the situation. With a like provision regarding other usages I will generally use the term hypnotizability for the capacity to develop or enter a state of hypnosis with and without reference to the so-called “depth” that it has presumably attained. Suggestibility will denote the capacity to produce what I have called the classical suggestion effect (Weitzenhoffer, 1974); that is, a nonvoluntary (or avolitional) response relevant to the content of a communication intended to be a suggestion. In contrast, I will also speak of hypnotic responsiveness and of responsiveness to suggestion in reference to the capacity to produce a response to a suggestion with and without hypnosis presumably being present. The reason for this distinction is that, in my opinion and experience, the observed responses are multidetermined, suggestibility being only one of their determinants. But it should be obvious that these scales of responsiveness must also (partially) measure, hence be scales of, suggestibility too.

Inasmuch as it is not possible to independently ascertain with any certainty that a person is hypnotized, that is, in a hypnotic state, I am prone to speak of hypnosis “presumably” being present. And again I am prone to speak of “so-called” hypnotic effects to clarify that I am speaking of effects that are conventionally said to be hypnotic even though the label is to be questioned. It would be more accurate to qualify them, in general, as being typical “suggested” effects with hypnosis presumably present. Likewise it seems probably more correct to speak of the possible influence of accepted hypnotic induction procedures than of hypnosis itself since the latter can only be assumed to have been brought about.

One last remark: My use of the word scale will frequently be a generic one, including under this label instruments used for assessing hypnotic behavior that technically speaking are not truly scales. Such are the Hypnotic Induction Profile (Spiegel & Spiegel, 1978) and the Stanford Profile Scales of Hypnotic Susceptibility, I and II (Weitzenhoffer & Hilgard, 1963)

Of Hypnosis Scales in General

To begin with, what is a scale as used in the present context? Briefly, it is an instrument, a device, so constructed as to provide a numerical assessment of various aspects of objects of interest to psychologists and other students of behavior, whereby some property of numbers can be associated with some property of the assessed attributes. In the present case, the objects of interest are certain features of so-called hypnotic behavior and the property is that of order or rank. The forerunner of modern, that is, post-1900, hypnotic scales was probably one developed by Hypolyte Bernheim (1884), but there definitely were much earlier attempts made at assessing certain aspects of this behavior (Weitzenhoffer, 1989).

There are a number of reasons why such a large number of scales exists. For one thing, many researchers and practitioners have had a special interest in being able to obtain an index of a person’s present or future potentials for producing hypnotic phenomena. This has led them to devise scales said to be of hypnotic susceptibility
and hypnotizability, that is, of proneness for developing these effects. Others have focused their attention upon a hypothesized hypnotic state presumed to be present and its presumed extent (degree, depth). For this they have devised scales of hypnotic depth, also confusingly said to be of hypnotizability, with the latter term more legitimately also often referring to simply the ability or capacity to develop a state of hypnosis regardless of its depth. Another motivation behind the devising of some of the scales has been the fact that early scales tested individuals by the simple expedient of directly checking how well they actually responded to induction of hypnosis procedures and to suggestions. For various reasons we need not consider at this time, some investigators and practitioners have sought to devise other, especially, indirect ways of doing this.

Still another reason is to be found in the fact that some of the early scales were, from a technical standpoint, deficient. Improving on them led to new scales, as in the case of the Stanford scales (Weitzenhoffer, 1997).

There also has been the time element. The early scales were time consuming. This led to devising shorter, if not simpler, procedures. One device, the Hypnotic Induction Profile, is vaunted by its author to take only five minutes.

Generally speaking the scales being considered were originally literally aimed at being hypnosis scales; that is, they were either intended to measure the extent (depth) or degree to which a person was in a hypnotic state, or their capacity to enter or develop such a state (hypnotizability). Although their names and the forms they take may at times suggest otherwise, this still appears to be one of the main reason for their existence. In any case, one obvious and seemingly reasonable way to accomplish the above that occurred to early researchers, was to sample the kinds of effects presumably hypnotized individuals could produce, that is, their hypnotic responsiveness. To do this, although their names have often not been consistent with it, a majority of the constructed scales have simply consisted of sets of test suggestions. At best, then, they are only scales of suggestibility, that is, of the capacity to respond to suggestions with a suggestion effect. They directly have had nothing to do with the hypnotic state.

There are a small number of other scales said to measure the depth of hypnosis that do not make direct use of suggestibility. Instead they are based on a sampling of how a person experiences being presumably hypnotized. However, the validity of most of these scales has been established on the basis of how well their measures correlated with measures of suggestibility obtained with other scales, particularly the Stanford Scales of Hypnotic Susceptibility: Forms A and B (Weitzenhoffer & Hilgard, 1959), or derivatives of these. Consequently, their acceptance as scales of depth has actually also depended upon their ability to indirectly give a measure of suggestibility. In this regard they do not differ from the first group of scales that have been discussed.

In brief, then, methods for assessing hypnotic depth, hypnotizability, etc., generally do not assess the state of hypnosis or the capacity to develop it. At best, they generally assess suggestibility and only suggestibility. There are a number of reasons for this having taken place. Although most workers, including this writer, who are interested in the state of hypnosis per se would agree this condition most likely cannot be reduced to merely being a state of suggestibility, (Bernheim, 1884; Weitzenhoffer, 1953; Evans, 2000) the fact remains that suggestibility has been the only feature intimately associated with it that can objectively be more or less numerically assessed.
Then, too, there has been a general feeling that suggested responses are somehow or other affected by the presence of hypnosis and reflect some sort of extent attributed to it. Finally, regardless of any connection suggestibility may have with hypnosis, there is little question that in today’s production and use of so-called hypnotic phenomena, including hypnosis, suggestion plays the dominant role in it. This last would particularly and naturally lead workers in the field to focus on suggestions and the capacity to respond to them in attempting to deal with hypnosis.

There have been a few proposed scales that do not fall among those just reviewed. With the exception of one we will later examine, most have seen little use. There is another, however, that possesses a rather interesting feature and which for this reason will also be later taken up.

It should be clear from the above discussion that the naming of the scales with which we are concerned does not always reflect the true character of the instruments and what they can or can not do (Weitzenhoffer, 1997). This has been a source of confusion and may have led to erroneous conclusions in some of the research done with regard to the hypnotic state. Readers should keep this in mind.

**What of Hypnosis?**

From the very start the existence of hypnosis, the state, has been more of an inference than an observed fact. Thus far, besides being one of the phenomena presumably brought about by suggestion, about the only thing one can say about hypnosis or, more correctly about existing so-called hypnosis inducing procedures is that, in some cases, they appear to cause suggestibility to become enhanced. Laboratory data have shown this enhancement exists, but also that it is much less than has generally been believed it can be. However, this last could be due to the highly restrictive ceilings imposed by existing scales of suggestibility that have been used, and the issue must remain an open one until further verification is done. This could probably be done by using the *Stanford Profile Scales of Hypnotic Susceptibility, I and II*. These were not available for the original testings of the enhancement hypothesis (Weitzenhoffer & Sjoberg, 1961). The importance of settling this issue will be made clear by the next paragraph.

This is not a well-recognized fact, but it has never been independently, empirically or theoretically shown that (a) hypnosis has or should have depth or degrees, and (b) that suggestibility is or should be a measure of this depth or degree. And since it is generally agreed that nearly everyone, if not everyone, is suggestible to some degree, hypnotized or not, a fundamental question is how in the first place can we differentiate between hypnotized and nonhypnotized individuals on the basis of suggestibility, that is, if this is possible. If it were simply a matter of the absolute amount of suggestibility present, no problem would exist. Unfortunately, no guidelines exist in this regard. The fact is that no simple or complex test or measure of suggestibility can alone even tell us whether a person is or is not hypnotized with any certainty. It cannot be overemphasized that, insofar as scales measuring suggestibility are concerned, anything one can say based on their measures regarding hypnosis must be understood to strictly be an inference and not a demonstrated fact.
More Caveats

As just pointed out, the majority of available hypnosis scales cannot directly tell us anything regarding the presence of hypnosis, its extent, or the subject’s capacity to develop it. Anything said in these regards is a pure guess, although possibly an educated one.

But if, when using so-called hypnosis scales, one cannot say with any certainty that one is directly assessing hypnosis or the capacity to develop it, it would seem that one can at least assert that suggestibility, or at least people’s responsiveness to suggestions, is being assessed.

There are two other features of the scale that tend to vitiate assessments made with them. For one thing, a necessary (although not sufficient) condition for a communication to be viewed as a suggestion is that the response to it be nonvoluntary, avolitional (Bernheim 1884; Weitzenhoffer, 1980, 2000). There is fairly broad agreement regarding this. But whether the response is or is not nonvoluntary is rarely obvious and, unfortunately, responses to communications intended to be suggestions can and do also take place at a volitional level. Since researchers and clinicians never check the matter, individuals frequently are credited with a suggestibility they did not actually demonstrate. In other words existing scales are widely open to false positive scoring. Only one group scale, The Carlton University Responsiveness to Suggestion Scale (Spanos et al., 1983) has built-in provisions for making the above check. However, this can also easily be done with any scales after the testing has been terminated.

There is also a problem regarding the extent to which scales presumed to measure one and the same thing do so and do so equally well. For more than forty years now the Stanford Scales of Hypnotic Susceptibility (Form A and B) have served as de facto standards for most other suggestibility scales and have served as a model for many. That is, most of the scales that were subsequently developed have been validated, some against these Stanford scales, but far more against other scales thus validated. All these scales have been considered to be equivalent to each other and, more particularly, to the Stanford Scales, hence interchangeable with each other, with the Stanford Scales and a satisfactory substitute for the latter. Unfortunately this equivalence and interchangeability may be more of a mirage than a reality. The reason is that the equivalence has been based on the existence of statistically significant correlations between the scales in question. The reasoning has been that if scale A is correlated with scale B, which itself is correlated with a Stanford Scale, then scale A is also correlated with this Stanford Scale. But this is not necessarily so! Furthermore, to add insult to injury, correlations that are far too low have been accepted as satisfactory (Weitzenhoffer, 2000). As a consequence, many of the existing scales said to be satisfactory substitutes for the Stanford Scales simply are not. This is not to say they do not measure something that is relevant, perhaps more so than what the Stanford Scales measure, or something that is useful in some other way, but what it is remains to be seen.

Going Beyond Suggestibility.

It was once believed that hypnosis, the state, created suggestibility. It was Lièbeault (1866) first, but especially Bernheim (1884) who convinced those working
with hypnotic effects that suggestibility exists independently of hypnosis. As the point has been made over and over in past pages, the two are quite distinct objects of thought and inquiry. It is therefore a legitimate question to ask whether one can assess hypnosis and the capacity to develop or enter it without assessing suggestibility or the responsiveness to suggestion. Some efforts have been made in this direction, most focusing on how individuals experience being presumably hypnotized. Unfortunately, as mentioned earlier, the construction of most of the instruments used for this purpose has been circularly guided by how well their scores correlate with suggestibility scores. That is, the experiential items that eventually became scale items have been selected so as to maximize these correlations. They may therefore constitute reasonably good indirect scales of suggestibility, but what they can tell us about the hypnotic state itself is a moot question. There are no good reasons for believing that scores measuring the hypnotic state or the capacity to develop it should necessarily also measure suggestibility or vice versa. It is also quite possible that hypnosis is an all-or-none phenomenon that has no depth or degree!

There does exist one hypnosis scale, or rather procedure, that attempts to directly assess the depth of hypnosis, and that does this without appeal to suggestibility. It also has the virtue of taking very little time for its administration. This is the subjective scale devised by Leslie M. LeCron (LeCron, 1953). Briefly, it calls for the presumably hypnotized person to make his or her own subjective assessment, using an imaginary yardstick for this, of how deeply hypnotized he or she is. Since, when this is done exactly as specified by LeCron, the resulting scores bear no relationship to suggestibility scores, there is some basis for feeling that one may be getting a true assessment of hypnotic depth (Weitzenhoffer, 2000). However, this case is lacking in solid evidence. One issue that has never been looked into is what it is the person being tested understands by the term “deeply,” or any variations of it thereof, when asked to numerically assess how deeply hypnotized they are! To what extent might this understanding vitiate the reports? The procedure has a certain appeal, being one that also allows one to easily make an ongoing assessment, from moment to moment, of the presumed depth of the hypnosis believed to be present.

One other instrument requiring very little time to use (five minutes, it is said) and that partially attempts to go beyond just measuring suggestibility is H. Spiegel’s Hypnotic Induction Profile (HIP) (Spiegel, 1972; Spiegel & Spiegel, 1978). I have discussed this instrument at some length elsewhere (Weitzenhoffer, 2000) and will only give a limited account of it here. Briefly, the HIP has two parts. First, a so-called Eye Roll test is administered and, then, a very brief test said to be of “dissociation.” The HIP is said not only to provide an assessment of hypnotizability but also to be an aid in the diagnosis of psychopathology. Thus far I have not been particularly impressed by the diagnostic usefulness of the instrument. In any event, lest there be some misunderstanding, whatever diagnostic capacity the HIP may have, its bearing upon the assessment of hypnotizability is at best unclear. A good point may be that the eye roll test portion of the HIP has been well shown to have little to do with suggestibility as measured by other instruments. But while H. Spiegel, who originated the test, has vaguely hinted (Spiegel, 1972) to an association between its results and the potentiality to develop a hypnotic state, he has never presented any sound grounds for supposing there was an association. Nor has he ever said how he came to make use of the eye roll test, and the latter seems to have been more or less in the nature of a shot in the dark.
The second part of the test does, however, correlate to some extent with suggestibility measures, but certainly not sufficiently to say it is a good substitute for a scale such as the Stanford scales as Spiegel and others have held (Spiegel, Aronson, Fleiss, & Haber, 1976; Frischholz, Spiegel, Tryon, & Fisher, 1981). This may actually be a point in its favor as a test of hypnotizability per se. Note, too, that Spiegel actually does not appear to consider the second part to be a measure of suggestibility, but instead seems to see it as a test of the capacity to dissociate. At best, however, it is a minimal test of dissociability, if at all one, and provides no measurement of it in any true sense of measuring. Also much depends here on what one understands by dissociation and how satisfactory the criteria used to ascertain its presence are. The Spiegels are not particularly explicit in this regard. Furthermore, if the HIP does assess such a capacity, it must be kept in mind that the notion hypnosis is a state of dissociation remains largely a hypothesis (Weitzenhoffer, 2000). A particular weakness of the instrument is a lack of provision for standardizing its administration. When demonstrating it, Herbert Spiegel has always used a fast paced, authoritarian, delivery. Apart for insuring a short administration there is nothing else known regarding how important this feature is. There is a likelihood many potential users of this instrument, left on their own, would administer it in a milder, slower fashion than Spiegel, particularly since the instrument contains no instructions in this regard. This could be an eventual source of problems regarding the utility of HIP data. In addition, the eye roll test is not satisfactorily standardized.

What and What Not to Do With the Scales

The scales as a research tool

Twentieth century researchers have been largely responsible for the development of so-called hypnosis scales. More specifically, we are talking of those doing basic research. However, the use of these scales also has a place in clinical research. Bernheim’s scale appears to have merely been the outcome of his efforts to describe the hypnotic condition. Some of the later scales came into being in order to better study individual differences in relation to the production of hypnotic phenomena. More generally, with the possibility of quantifying suggestibility, and possibly hypnotic depth, the door has been open to relating such quantification to a large variety of other quantifications. More accurate replications have become possible. Research calling for an accurate preselection of suggestibility levels has greatly benefited too. The group scales such as the Harvard Group Scale of Hypnotic Susceptibility (Shor & Orne, 1962) are particularly useful where there is a need to maintain a pool of subjects of varying suggestibility and when one needs to sort a large sample of individuals in this regard.

In so far as measuring suggestibility is concerned, the most useful scales are, in my opinion, the Stanford Scale of Hypnotic Susceptibility, Form C (Weitzenhoffer & Hilgard, 1962) and K. Bowers’ modification of it for group testing (Bowers, 1993). For one thing, in contrast to Forms A and B and various derivatives of these, suggestibility is the only focus of Form C. Because it is a Guttman type scale it is not necessary to administer the entire scale to low scoring individuals, thus saving testing time. It also does a better job of picking high scoring subjects than Forms A and B. Unfortunately, any of the Stanford Scales (except Bower’s modification) are currently
hard to obtain, having been allowed to go out of print. There are, however, indications they may again become readily available.

One should keep in mind the caveats listed earlier that apply to all scales of suggestibility. In particular, they cannot tell us with any definitude that a person is hypnotized or how deeply. Because researchers have failed to recognize this, I believe that much of the research done specifically with respect to the state of hypnosis is equivocal. Many of the subjects that have been assumed hypnotized have most likely been only suggestible. Any conclusions made regarding the state of hypnosis from data collected with them more likely pertains to their suggestibility. I particularly have in mind here physiological data that have been reported. These may have had little to do with hypnosis. As a rule of thumb, I can only offer my feeling here that a score of at least 10 on the Stanford Form C may justify speaking of hypnosis being present. A much safer approach is also to use the Stanford Profile Scales of Hypnotic Susceptibility, I and II, preferably both. A superior performance on these most likely indicates the subject is in a hypnotic state, the more so that one knows for a fact the effects demonstrated with the Profile Scales are not a part of the subject’s habitual behavior. Admittedly the use of these scales is time consuming, but for critical research this may be time well spent. However, the only sure way for the time being remains that of showing an appreciable increase in responsiveness or suggestibility has resulted from the use of the induction of hypnosis procedure that was employed (Weitzenhoffer, 1980, 2000), particularly if the Profile Scales have been used. Although this should not be a deterrent to the use of the various existing scales and will not apply to many uses of the latter, it should also be kept in mind that, to date, all quantifications of suggestibility and depth have been of a relatively primitive kind, being at best ordinal in nature. This last feature places limitations on the interpretation and mathematical use of the scores that are obtained, as well as on the kind of statistical work that can be done with them (Weitzenhoffer, 2000). These limitations have too often been disregarded in past research.

The Scales as a Clinical Tool

Various objections have been voiced by clinicians regarding the use of the scales discussed in this article. As I have explained in detail elsewhere (Weitzenhoffer, 2000), the only clearly valid one at present is the time involved in administering them. Some of the other objections may be valid, but this needs to be demonstrated, not just hypothesized.

There are psychologists and psychiatrists who feel that every prospective psychotherapy patient should be routinely widely assessed and that this should include assessments of hypnotizability and the like. I have seriously questioned any practice of routinely including tests of hypnotic responsiveness in the test batteries that may be used. I have yet to hear convincing reasons for doing so. Any testing should be guided by a specific need for it. In this regard, many practitioners seem to feel that if they know ahead of time how suggestible their patient is they will be better able to plan the therapy. I agree that if one plans to use a specific hypnotic effect one needs first to know whether it can be produced. There is no better way to do this than to go ahead and produce the effect, but this does not entail using a whole battery of test suggestions, as so many therapists envisage doing. Furthermore, most available scales
only test, at best, for the ability to produce a dozen or so effects and the one of interest will often not be among those produced by the scale that is used. True, sometimes there may be a scale item that comes close enough to the effect one has in mind that one feels if the patient passes it she will also be able to produce the desired effect. Most often this does not happen. On the other hand, if it should happen that the desired effect is a test item, how useful is it to know ahead of time what other effects the patient can produce? How useful is it to know, for instance, that a patient can do an age regression when all one needs to do is to produce an analgesia? Generally, not greatly. For this reason, in my own practice of clinical hypnotism I do essentially no pretesting of suggestibility. My advice is not to spend time on testing, which can be better spent doing something more constructive.

Of course, if one’s clinical work is also part of a research project that is a different matter. Also, I do believe that there is something to be scientifically gained if a therapist provides a reliable measure of the patient’s overall level of suggestibility as measured by an accepted scale even though there is no ongoing research on his part. For one thing, someone may come along later who may be able to use this information in a survey of like cases. It may also be useful to another clinician who has a like case and is searching through the literature for treatment ideas. Such information may even be useful for individuals doing basic research. So while I am saying that for one’s own practice of clinical hypnotism preliminary assessments with scales are not a requisite, I am also advocating that it can be of value at least to others, and I would encourage every clinician using hypnotic techniques to at least give his estimation of where on a scale, such as those of Stanford, the patient might fall. For this reason I also strongly recommend that all clinicians using hypnotic suggestions familiarize themselves with the Stanford Scales or a derivative of these. When reporting suggestibility levels they should also specify what scale they have in mind.

The caveats mentioned earlier should of course be kept in mind by clinicians using any of the scales. However, while the issue of the scales being doubtful instruments for the determination of the presence of hypnosis or of hypnotizability is of concern for researchers, it should not be for clinicians otherwise using the scales. For, after all, what is important in clinical work is not whether the patient is hypnotized but how suggestible he or she is. If he or she is sufficiently suggestible for the purpose of producing the needed effect, this is all that counts. Whether or not a state of hypnosis is present is not the issue. From this standpoint, Form C of the Stanford Scales is probably as good a scale as any to use in this context.

I hear complaints on the part of clinicians that the scales do not give information relevant to the clinical outcome of the use of hypnotism. True, but it needs to be kept in mind that the scales were not designed with that aim in mind. Undoubtedly, devices capable of doing so would be useful and this might be something research oriented clinicians might consider doing in the future rather than constructing more scales of responsiveness and depth.

References


