Psicothema 2021, Vol. 33, No. 3, 378-385
doi: 10.7334/psicothema2021.24

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Seven Decades of Factor Analysis: From Yela to the Present Day

Pere J. Ferrando
Universitat Rovira i Virgili

Abstract

Background: I review some key developments in the exploratory factor analysis (EFA) model from the 1940s to the present day with Yela as the reference point. **Method:** The study is organized in three parts. The first part (until 1950) discusses EFA during Yela's formative years. The second part reviews the evolution of the model from Yela's return to Spain to the end of the century: the development of factor analysis (FA) as a statistical method, and the advent, and unstoppable rise, of the confirmatory model. An interlude follows in which some key problems are discussed. The third part discusses the resurgence of the model in the 21st century and the advances made in this period. I end by discussing the current situation in the field. **Discussion:** I discuss the validity of Yela's views regarding FA, and criticise how technical and instrumental developments have favoured, in most cases, mindless applications of the technique in which the substantive and design aspects of the research are grossly neglected. **Conclusions:** Although new technical developments are still needed, EFA is finally at par with any structural model. So, its time again to give importance to what is really important: the design and the substantive aspects.

Keywords: Mariano Yela; exploratory factor analysis; semi-confirmatory factor analysis; psychometric applications.

Resumen

Siete Décadas de Análisis Factorial: Desde Yela Hasta Hoy**. Antecedentes:** se revisan algunos desarrollos clave en el modelo de análisis factorial exploratorio (AFE) desde los años 40 del hasta hoy, tomando a la figura de Yela como punto de referencia. **Método:** el estudio se organiza en tres partes. La primera (hasta 1950) revisa la situación del AFE durante los años de formación de Yela. La segunda revisa la evolución del modelo desde la vuelta de Yela a España hasta el final de siglo: el desarrollo del AF como método estadístico y el inicio e imparable ascensión del modelo confirmatorio (AFC). A esto sigue un entreacto en el que se discuten algunos puntos clave. La tercera parte revisa el resurgir y los avances del AFE en el siglo 21. Se termina discutiendo la situación actual. **Discusión:** se defiende la validez del punto de vista de Yela y se critica que los desarrollos técnicos hayan favorecido aplicaciones poco fundamentadas en las que los aspectos sustantivos y de diseño se encuentran inaceptablemente descuidados. **Conclusiones:** aunque siguen siendo necesarias mejoras de tipo técnico, el AFE está ya a la par con cualquier modelo estructural. Es hora de volver a dar importancia a lo que realmente la tiene: el diseño y los aspectos sustantivos.

Palabras clave: Mariano Yela; análisis factorial exploratorio; análisis factorial confirmatorio; aplicaciones psicométricas.

Factor analysis (FA) plays a central role in Mariano Yela's remarkable work. First, as a student and young researcher in the late 1940s he worked with the leading experts in the field (particularly Thurstone; see Yela, 1996a), becoming a skilled FA researcher. Once back in his home country, he introduced the technique into Spain and taught several generations of researchers and practitioners (Muñiz, 1995, 2021; Pascual et al., 1997). Finally, most of his own empirical research was based on FA. So, reviewing the evolution of FA from the late 1940s to the present day by taking Yela as a reference point seems well justified. It is, however, a more complex task than it might first appear.

One major initial problem is the wide scope of the content: FA has expanded considerably both at the methodological and the applied level, far outstripping the psychological domains from which it was created (e.g. Ferrando & Lorenzo-Seva, 2017). To solve this problem, I decided to limit the review to (a) unrestricted or exploratory FA (EFA; I will use both terms indistinctly, although unrestricted is far more appropriate) that Yela focused on, and (b) the psychometric applications of FA that were the focus of Yela's research and interests. I am reasonably sure that this second decision is correct. I had the pleasure to introduce Yela at a conference, and he explicitly defined himself as a psychometrician.

Starting with Wolfle (1940), excellent reviews on the evolution of FA have appeared from time to time in the literature (Mulaik, 1986; Cudeck & MacCallum, 2007; Mislevy, 1986). Good didactic presentations of the present state of the technique are also available in our country (Izquierdo et al., 2017; Lloret-Segura et al., 2014). So, a new, purely objective, review would not make much sense. Instead, I decided to base the review on my own personal standpoint and some of the remarks that Yela made on multiple occasions, most of which I fully agree with. Some readers might think that certain points that I stress here are not so relevant, while I do not even consider more important points. They might not agree with my position either, which is perfectly fine. My review is also very critical: I believe that the computational, methodological, and instrumental developments in FA have encouraged mindless applications of the technique in which the substantive and design

Received: January 21, 2021 • Accepted: March 22, 2021
Corresponding author: Pere J. Ferrando
Universitat Rovira i Virgili
43007 Tarragona (Spain)
e-mail: perejoan.ferrando@urv.cat

aspects of the research are grossly neglected. Omitting criticisms of this type would do a disservice to Yela's work: he repeatedly insisted that the substantive bases and the research plan were the most important and difficult aspects of FA research, while the technical aspects were, by comparison, secondary and relatively easy (Muñiz, 1995; Yela, 1996b).

Restrictions aside, the review is admittedly too simplistic. The relevant literature I reviewed is rich in nuances, changes of opinion, or plain contradictions (to which I have added mine). Having made this caveat, the rest of the paper is organized as follows. First, I review the evolution of FA from the 1940s until Yela returned to Spain. Second, I review the period between 1950 and the end of the 20th century, during which time FA became a statistical method and CFA the dominant paradigm. I then take a short break during which I discuss some key points which are needed in order to assess whether the dominance of CFA is justified. Finally, I end by discussing the current situation in the field and attempting to make some recommendations.

A step back in time: FA during Yela's formative years

In the small world of FA, the 1940s were exciting, controversial years. The most visible trend was that the group from Chicago –the Thurstonians as Mulaik (1986) refers to them– were clearly taking the lead over the group from London and imposing the correlated-factors model over the British hierarchical models derived from Spearman's theory. However, things were far more complex than that in this period. Also at Chicago University, Holzinger was developing the bifactor model in the purest British tradition. And other groups were making other important contributions. In particular, in Edinburgh, Lawley (1940, 1944) was working on two theoretical developments that would revolutionize the field 20 or 30 years later: The statistical estimation of FA by maximum likelihood (ML) and the FA of categorical variables.

Although his training was much broader, the period in which Yela was mainly influenced by FA was between 1946 and 1948 when he was studying with Thurstone himself (Yela, 1996a). So it seems pertinent to start by reviewing the direction and aims of this group in order to get a (partial) idea of how things were at that time. Some facts may be a bit surprising given present views on FA.

The research carried out by the Thurstonians can clearly be qualified as confirmatory (Thurstone never believed in purely exploratory FA). So, they were actually using the unrestricted (exploratory) FA model to perform confirmatory-oriented studies. More specifically, the typical research approach in this group involved (a) defining hypotheses about the number and structure of the factors, (b) selecting measures expected to reflect the hypothesized factors, and (c) assessing the extent to which the rotated final solution was in agreement with the initial hypotheses. In this schema, it is worth taking a closer look at stage (b): the designs and data-collecting plans were well thought out, and great care was taken to ensure that the factors were overdetermined, with multiple indicators each, that the measures were as pure as possible so that the simple structure was easier to obtain, and that they were also free from redundancies or shared specificities that could give rise to correlated residuals or 'bloated specific' factors. This way of working could perfectly be used as a guide today, and would also reveal the flaws and weaknesses of many factorial studies at present.

The main limitation of this approach was that, at this time, unrestricted FA was still not in a position to rigorously assess either the dimensional (i.e. number of factors) or the structural (i.e expected structure) hypotheses required in confirmatory research. Thurstone assessed them by eye, but was well aware that this was a limitation. Even so, the methodological research of the group never went in this direction: Thurstone regarded FA as a mathematical method, and testing the hypotheses above would have required taking a statistical, inferential view. In fact, Thurstone was explicitly against a statistical view of FA, because he considered that random sampling was inappropriate in this type of analysis, and that the best sample consisted of extreme examples of the factors under study (Butler, 1968).

Overall, the contributions that the group had made by the end of the 1940s were impressive. In relatively few years, Thurstone developed (a) the equations of the multiple FA model, (b) the concepts of communality, uniqueness and the reduced correlation matrix, (c) oblique rotation and the concept of simple structure and (d) residual analysis in FA (see e.g., Bock, 2007; Wolflee, 1940; & Thurstone, 1947). In contrast, the purely technical achievements are, in my view, largely of historical interest. In particular, the rotation procedures proposed seem unnecessarily complex by today's standards: rotation can be solved in much simpler ways (e.g., Browne, 2001; Lorenzo-Seva, 1999).

Judging by his 1949 *Psychometrika* paper (Yela, 1949), Yela was at that time a technically skilled researcher, but much more importantly an independent thinker, with great curiosity, who did not comply (at least not all of the time) with the demands of his group. To start with, his initial interest was to reconcile the British and the American views, while Thurstonians tended to ignore or dismiss the Spearman-based hierarchical modelling. Second, he did not believe in a confirmatory approach but rather thought that the task of FA was to **discover** structures rather than imposing them a priori: the eventual confirmation of these structures would be then given by the replicability across different samples, methods or measures (García-Jiménez & Yela, 1995; it seems that he continued to think along these lines until the end). Finally, he made a far more thoughtful and deeper interpretation of the results than was usual in FA applications of the time.

If I had to assess everything that Yela brought with him on his return to Spain, I would first consider the key methodological contributions summarized above. But perhaps of even greater significance was the fact that he attached importance to what really mattered: the care that should be taken in the design and substantive basis of the study, and in the interpretation of the results. This distinctive attitude is in stark contrast with the trivialization that technical advances have generally brought. Finally, I should mention two basic Thurstonian principles that, in my view, are timeless and which should serve as a guide in any FA study (Thurstone, 1947): Firstly, that the art of FA is mainly to consider a reasonable minimal number of factors compatible with sufficiently low residuals and, secondly, that inspection of the residuals is the soundest way of judging the appropriateness of an FA solution.

Times are changing (1950-2000)

After various research visits abroad, Yela returned permanently to Spain around 1952 (Yela, 1996a). Far more authoritative voices than mine have discussed the enormous repercussions this decision had on Spanish Psychology, both in this volume (Muñiz, 2021) and in previous papers (Muñiz, 1995; Pascual et al., 1997; Siguan, 1994). As far as this review is concerned however, Yela became a teacher and a researcher who used the technique but did not directly contribute to the advances in this period.

The two decades that followed Yela's return witnessed another revolution in the small world of FA. However, it was of a very different type to the previous one. The bases of the controversies in the 1940s and the changes were theoretical and substantive. Those of the ensuing decades were methodological.

The main change that started in the 1950s was that FA gradually moved away from being seen as a mathematical and descriptive method to become more and more an inferential statistical method. In turn, the inferential approach took two main directions (e.g. Butler, 1968; Cliff & Hamburger, 1967; Jöreskog, 2007). In the first, conventional statistical approach individuals were regarded as a sample from a hypothetical population, while the measures (tests or item scores) were considered to be fixed. In the second, which is usually known as psychometric inference, tests or items were viewed as samples of measures from a large domain defined by all the hypothetical measures that could be developed for measuring the dimensions of interest. It now seems clear that statistical inference won the battle. However, some of the most original ideas in FA were developed from psychometric inference, and I do plan to focus some attention on it. In particular, I will focus on Guttman's (1955) image analysis.

Guttman's initial interest was to avoid indeterminacy in defining the common and unique parts of the FA model. However, this indeterminacy is closely related to the most fundamental indeterminacy in FA (and usually swept under the rug): that a given structural FA solution is compatible with an infinite set of different individual scores on its factors (i.e. univocal measurement cannot be warranted from an FA model).

The approach taken by Guttman to avoid indeterminacy was to define the common parts of the variables as the parts that can be predicted from the remaining variables in the domain by multiple regression, so the correlation matrix between these common parts (the image matrix) must be the basis of the FA. Note that Guttman's definition refers to the population: it holds when the infinite set of possible measures that define the domain is considered. The image obtained with a given number of measures is a sample estimate.

The insights derived from this approach are now summarized: As the number of variables that define each factor increases, and as the simplicity and accuracy of these variables improve, the image matrix gets closer and closer to the conventional reduced correlation matrix with proper communalities on the diagonal. And, more importantly, the infinite sets of possible scores compatible with the solution are more and more similar to each other, up to the point that indeterminacy is nearly removed (Bollen, 2002; Ferrando & Lorenzo-Seva, 2018). For this to occur however, the indicators should be truly representative of the domain they aim to measure (Mulaik, 2010).

So, where does this lead us? It seems clear. To reduce the ugly flaws of indeterminacy, the variables must be carefully chosen to represent the domain they attempt to measure, they must be as factorially simple as possible, and they must have a small amount of measurement error. And, above all, each factor must be overdetermined, and defined by a sufficient number of indicators. So, the solution lies in the design and not in the mathematical formulation, and the principles of the design are those set by Thurstone, to which Yela adhered all his life.

We turn now to the statistical-inferential approach. As is usually the case, this approach was finally made to work by the early contributions. The equations for ML estimation of the factor model had already been provided by Lawley in the early 1940s. However, since the solution is iterative, both (a) an efficient algorithm and (b) enough computational power were still needed to make it feasible (see e.g. Lawley & Maxwell, 1963). Howe and Schönemann (see Mulaik, 1986) made initial attempts, but it was Jöreskog (1967) who developed the general approach that, with improvements and refinements, has lasted to this day. This approach provided not only ML estimates of the structural parameters, but also an inferential test of fit for determining the number of common factors.

I shall discuss the main advantages of the statistical approach below. However, an important point noted by Jöreskog (2007) and which concerns the change of direction in FA may be appropriate here. Up to the 1960s, while FA was considered to be mainly a mathematical method, it was developed by psychologists (with a strong mathematical basis). For better or for worse, the statistical developments since then have been led by mathematical statisticians.

The development of FA as a statistical model was the key that opened a Pandora's box containing both blessings and evils. Once the hypothesis regarding the proper number of common factors was testable, only the hypothesis about the closeness to simple structure remained to be worked out. Bargmann (1957) had proposed an initial but intricate procedure (which Yela, 1968, explained in a far clearer way). However, Jöreskog (1969) realized that a much simpler test could be obtained by (a) specifying a priori as many parameters of the solution as possible, (b) fitting the restricted pattern with ML and (c) assessing the appropriateness of the proposed solution with the associated test of fit (note, however, that this test now jointly assesses two hypotheses: the correct number of factors and the relational structure). Confirmatory factor analysis (CFA) was born.

As is well known, since the early 1970s, the focus of both methodological and applied interest in the field of FA changed from EFA to CFA. At the methodological level, developments in EFA slowed down and finally almost stopped. And, at the applied level, it was relegated to a minor role, to the point that it was viewed (at best) only as a rough preliminary step for "cleaning" the data before a "proper" CFA could be fitted to it (e.g. Bollen, 2002). Ironically, this change took place in a period in which EFA was advancing in leaps and bounds: meaningful, powerful analytical rotation procedures that were also simple and clean were being proposed (Browne, 2001); semi-confirmatory solutions based on target rotations were being rigorously developed and work was even started on goodness-of-fit indices that would be used much later in structural equation modelling (Tucker & Lewis, 1973). In my opinion, it is doubtful that this abrupt change in priorities was a step in the right direction.

From a purely modelling point of view, there is no convincing reason to favour one type of FA over the other. In spite of the efforts many authors have made to emphasize their differences, EFA and CFA are not really different models, but different types of solutions within a common model that differ in the structural constraints they impose. This point has been repeatedly made in the literature (see Ferrando & Lorenzo-Seva, 2000 for a review) but even so, it does not seem to have had much impact. As a practical exercise I use in FA courses, I show that a rotated EFA solution can be fitted by standard Structural Equation Modelling (SEM) programs (such as LISREL, Mplus, or EQS) by specifying it as if it was a CFA solution with minimal constraints. Students are generally convinced that they are not dealing with different models.

Given the discussion above, it seems clear that virtually all the improvements that have taken place since Jöreskog's initial proposals could have been applied to both EFA and CFA. So both types of solution could have evolved at the same time, could have been applied on equal grounds in substantive research, or could even have been combined in the same analysis. An interesting study by Jöreskog (1969), for example, obtains a mixed correlated-factor solution in which some factors are specified by CFA constraints whereas others are determined by EFA rotation (Carlson & Mulaik, 1993, made a similar combination). Nowadays, this "hybrid" use of EFA and CFA solutions is implemented (and so potentially feasible) in ESEM (Asparouhov & Muthén, 2009), but it would probably be hard to swallow for the dogmatic, rigid reviewers who are so common nowadays.

The main advantage that CFA has over EFA is that, because it is far more constrained, it is simpler, easier to estimate, more determinate, and potentially more stable. On the other hand, it is far less flexible (Ferrando & Lorenzo-Seva, 2000; Gerbing & Hamilton, 1996), something that tends to be forgotten. Again, can the choice of CFA **always** be justified by invoking the parsimony principle? As I see it, responding to this question requires us to consider some basic issues, which I discuss in the section below.

Intermission: Some basic issues and considerations

A review of the literature up to the 1970s clearly shows that the usual units of analysis in psychological EFA applications were test scores. In fact, EFA was explicitly considered to be a method designed to analyse a battery of tests (Tucker, 1983). When this is the case, the number of variables will necessarily be small to medium: obtaining a battery of, say 15 to 20 full-test scores is extremely demanding. Second, test scores can be made to be both highly reliable and considerably 'clean' and simple in the FA sense (e.g. Guilford, 1952). So, if it is designed well, the EFA of a battery of test scores is expected to provide a strong, replicable solution that approaches simple structure.

In the development of psychometric EFA applications, however, the units of analysis were changing more and more from test scores to item scores. In fact, EFA is viewed at present as a basic tool for assessing the dimensionality and internal structure of a set of items, which is indeed a key assessment in the process of test construction (Muñiz & Fonseca-Pedrero, 2019). So, it is instructive by contrast, to study how leading factorialists such as Guilford, Cattell, Eysenck, and Comrey, who were the first to make the test-item transition, developed their multidimensional measures. The impression one gets is that they took great pains to avoid the direct FA of their item banks. Common strategies were: (a) to analyse the bank in separate parts (e.g. on a scale-by-scale basis), (b) to undertake massive 'cleaning' and reduction of the initial bank by using conventional item analysis techniques, and/or (c) to group the items in parcels (facets, testlets, etc.) and use these parcels as units in the FA. Apart from the computational limitations of the time, there were good reasons for these strategies. To start with, an item bank is generally far bigger than a battery of tests. But, in addition, item scores are inherently unreliable and complex in the FA sense (e.g. Cattell, 1986). As a result, it is difficult to obtain strong factor structures on the basis of large sets of item scores (Tucker, 1983). Rather, solutions tend to be weak, complex, and difficult to replicate. And I want to point out that I am still talking about exploratory FA here.

While Jöreskog's work on CFA was methodologically a breakthrough, at the applied level he continued with the test-score-unit tradition: most early CFA applications consisted of analyses (or better re-analyses) of battery scores that came from good designs. In these conditions a CFA solution was expected to work quite well. Indeed, it was expected that most cross-loadings fixed to zero were not exactly zero, just rather low. And, if the number of variables was not too large, the misspecifications would be tolerable (see Ferrando & Lorenzo-Seva, 2000). To sum up: If a well-designed battery of test scores is to be factor analysed, a CFA solution is likely the best choice there is.

For the reasons discussed above, however, trying to fit a CFA solution to a large bank of item scores is an invitation to disaster or, as the saying goes, it would be like trying to force square pegs into round holes. In my opinion, the FA in this case should almost always be unrestricted and the appropriateness of the solution should be judged by criteria that go beyond pure model-data fit (as I discuss below).

It should be clear by now that the idea that EFA should replace CFA and again become the dominant paradigm is, in my view, as ill-founded as the present dominant position that CFA is the only proper way to analyse **any** set of data. Different problems and different datasets require different tools, and skilled researchers should know which is the most appropriate for their research. What is more, as I discuss below, if the artificial EFA-CFA distinction is abandoned, the researcher will find a continuum of possibilities ranging from a totally restricted solution to a fully unrestricted solution.

The distinction between test scores and item scores is a good illustration of how the nature of the data should guide the choice of the FA solution and method. So, I will continue along these lines. The choice will also depend on the metric properties of the scores. Test scores generally approach the continuity and linearity assumptions of the standard linear model. Item scores are clearly ordered-categorical and in principle, a non-linear model that treats them explicitly as discrete and bounded variables is theoretically more appropriate. However, things are not so simple. I have repeatedly discussed this issue elsewhere (Ferrando & Lorenzo-Seva, 2014) and I shall not discuss it again here.

A less common issue that I prefer to discuss instead concerns the ultimate aims of the analysis. When the units are test scores, then the main aim of FA is indeed to assess the dimensionality and relational structure among these scores: after all, the tests have already been scored. However, when FA is used for purposes of item analysis, the assessments of dimensionality and structure can only be considered as intermediate aims: what FA is doing now is calibrating a set of items that will form a test (Muñiz & Fonseca-Pedrero, 2019), and the ultimate aim of a test is some form of, generally individual, measurement (such as assessment, screening, classification, selection, or change). If FA is indeed used in this way, the accuracy and validity of the individual factor score estimates derived from the calibration stage should be the most important properties required for assessing the FA model.

Putting all this together, it should be clear that rigid "universal" positions are unacceptable. Thus, CFA will generally be the best choice for analysing a battery of tests, and the appropriateness of the solution should be mostly based on model-data fit assessment.

EFA, on the other hand, will be far more suitable for analysing large sets of item scores with complex and relatively weak structures, and the properties of the scores derived from the analysis may be more relevant measures of appropriateness than pure goodness-of-fit assessment.

Modern times: Where are we now?

From the very beginnings of CFA dominance, many practitioners have been aware that a restricted solution was generally too parsimonious for the complexity of their data (e.g. McCrae et al., 1996). However, until well into the 2000s, they felt that they were preaching in the wilderness (see Ferrando & Lorenzo-Seva, 2000). Once the tide started to turn, however, developments in unrestricted FA were numerous and rapid, and, at present, the gap between EFA and CFA is finally closing. I would stress again, however, that the turn of the tide seems to have taken place only in methodological forums. At the applied level, CFA continues to be the preferred and inevitable choice: some inertias are hard to change.

A summary of the developments in EFA in the 21st century would make for a complete article. So, my summary here will necessarily be incomplete. I will review improvements at two levels: internal and extended (to external variables or to multiple groups).

The internal level refers to improvements within the basic EFA solution. The initial issue here is that of the estimation and testing procedures, and the summary is clear: EFA can be fitted and assessed in the same ways as any structural equation model, and most of these ways have already been implemented in widely used packages (see Browne, 2001; & Ferrando & Lorenzo-Seva, 2017). Some reflection is required, however, to appraise the relevance of these advances. If interest is mostly on the point estimates of the structural parameters, and, (a) the design is well thought out, (b) the sample is reasonably large, and (c) the structure is strong, then the practitioner will likely be disappointed, because the estimates obtained from using the more sophisticated procedures now available would be almost identical to those provided by the previous, far simpler approaches, something that Yela was well aware of (Yela et al., 1969; García-Alcañiz & Yela, 1980). This is because the theoretically superior procedures use more available information from the data and are more efficient. However, virtually all the available procedures in EFA are consistent, which means that, if the model is reasonably correct, the different estimates will all converge towards the same values as the sample size increases.

As I see it, the main advantage of statistically based estimation procedures such as ML or weighted least squares (WLS) is that they allow standard errors of the parameter estimates to be obtained analytically, and this is a substantial step forward in EFA usage. Standard errors allow confidence intervals for loadings or inter-factor correlations to be obtained, and these in turn allow important decisions to be rigorously made about such things as the salience of a loading or the convenience of an orthogonal rotation. Having said that, our choice in FACTOR (see Ferrando & Lorenzo-Seva, 2017) was to obtain standard errors and confidence intervals for all EFA estimates (including factor scores, reliabilities, and goodness-of-fit indices) using intensive re-sampling (Bootstrap) procedures. In this way, this important piece of information can also be obtained by using simpler estimation procedures (particularly ULS), which, in our experience, are more robust and stable with large datasets (particularly for categorical variables).

In addition to standard errors, statistical estimation of the FA models allows a rigorous model-data-fit assessment to be performed, which theoretically avoids the use of approximate procedures and rules of thumb that have been the norm in EFA for many years. The initial proposals (Jöreskog, 1967; Lawely & Maxwell, 1963) consisted of developing a likelihood-ratio statistic that approached a chi-square distribution if (a) some key assumptions were met, and (b) the proposed model was correct in the population. This approach, however, had two main shortcomings. First, if the preliminary assumptions are not met (mainly that the variables are continuous and multinormally distributed), the distribution of the statistic can greatly depart from the reference distribution. Second, and more importantly, no FA solution is expected to be exactly correct in a population: rather, models are expected to be reasonable approximations. So, the null hypothesis of the test is always false, and will be unavoidably rejected as soon as enough power has been attained. The issue then is not to test whether the proposed solution is correct or not but rather whether it is a close enough approximation.

Goodness-of-fit (GOF) SEM-based research has greatly evolved since the naïve initial proposals above. As for the first limitation, robust corrections of the test statistic have been developed that closely approximate it to the reference distribution even when distributional assumptions are not met, or the efficiency of the estimation procedure is less than expected. As for the second, numerous indices address different facets of fit and allow close-fit tests to be performed and measures of approximation error to be obtained.

In spite of the valuable advances above, however, I believe that statistically-based GOF assessment is not the same as appropriateness assessment, but only a necessary condition. To explain this point, I would note that a weak and unstable solution, unlikely to be replicated in new samples, that provides indeterminate and unreliable scores can be compatible with acceptable fits in pure GOF terms. Many other authors have recognised this point and made proposals that go beyond pure model-data fit (e.g. Rodríguez, Reise & Haviland, 2016). In the specific field of EFA, we have developed a comprehensive proposal of this type that aims to assess (a) the strength, quality and replicability of the solution, and (b) the interpretability, accuracy and determinacy of the factor score estimates derived from it (Ferrando & Lorenzo-Seva, 2018). So, an acceptable EFA solution, must not only fit acceptably in GOF terms, but must also be strong, clear and replicable, Furthermore, if the ultimate aim is individual measurement, it has to provide reliable and determinate score estimates.

Still within the basic EFA solution I would like to mention two important improvements. First, improved estimates of the factor scores that (finally!) include standard errors and that use more information from the data (or prior, if appropriate) are now available (Ferrando & Lorenzo-Seva, 2016). Second, the procedures for transforming the initial solution (i.e. rotations) have evolved considerably, and there is now a range of choices that go from purely analytic rotations to target rotations with different levels of specification (Browne, 2001). Hybrid solutions that perform a target rotation in which the target is initially derived analytically are also available (Lorenzo-Seva, 1999). Standard errors for the rotated loadings are available for any rotation procedure, and congruence and fit measures for the solution obtained in the target rotations are also available (Lorenzo-Seva & Ferrando, 2020). Today the flexibility of unrestricted FA as a semi-confirmatory method is considerable.

Finally, I would like to consider extensions in which the "internal" EFA solution is embedded into a full SEM or generalized to multiple groups. To judge whether these extensions are really relevant, I would like to summarize the proposals that Yela made at the end of his career regarding the structure of intelligence (e.g. Muñiz & Yela, 1982; Yela, 1987). The proposal was an oblique hierarchical structure, with higher order factors that ended in a general factor. However, this structure is expected to vary in groups defined by age or cultural level. In some cases this variation will essentially maintain a common solution with different dominances. In others, the structure itself might change in the form of an increasing number of factors that are progressively differentiating. In any case, the structures obtained only reach psychological relevance if (a) they systematically appear across studies, and (b) can be validated via relationships to relevant external variables.

Now, let us translate these ideas to methodological terms, and consider what an FA solution would be required to accomplish in order to assess Yela's substantive model. Clearly, we should be able to (a) estimate and fit a higher-order EFA model at more than two levels. Furthermore, (b) this solution should be fitted simultaneously in different populations, and different levels of invariance should be assessed. Finally, the idea that the rotated solution acquires most of its meaning when related to extra-factor variables implies (c) extending the measurement model to a full structural model (which includes criteria or relevant external variables).

Extensions (a), (b) and (c) are fully feasible within a CFA approach, and this would be my recommendation for a study of this type: analyses should be based on test scores, and both the internal and external measures should be carefully chosen. However, we are dealing with EFA here. So, let us assume that we have to analyse a complex measurement structure (probably based on item scores). In principle, higher-order factoring which includes all the internal advances considered above is perfectly feasible within an EFA framework (Ferrando & Lorenzo-Seva, 2017). Multiple-group analyses, and testing for different forms of invariance/differentiation within an EFA solution is however a more complex issue. In theory, it can be performed via ESEM (Asparouhov & Muthén, 2009). It can also be approached by rotating the different solutions to a common solution that has both maximum simplicity and agreement, and assessing the degree of agreement (Lorenzo-Seva, Kiers, & ten Berge, 2005). However, I believe that this issue is more naturally addressed within a CFA solution, and that further research for the EFA case is warranted. Finally, the full structural analysis based on the core EFA solution can be performed directly via ESEM (i.e. simultaneously estimating all the structural parameters) or indirectly via extension analysis (e.g. Devlieger, & Rosseel, 2017). In this last case, (a) factor score estimates are obtained from the EFA measurement part, (b) external variables are regressed on these estimates, and (c) the estimated structural regression parameters are corrected for attenuation. Because it is more complex but uses less information, this last approach has been traditionally considered inferior to the joint structural estimation. However it has at least two important advantages. First, it can be used even when the full model is not identified or estimation becomes intractable or unfeasible in practice. Second, if the EFA core part has specification errors, these errors will propagate much less to the structural part of the model (e.g. Gustafsson & Balke, 1993).

Discussion

At the methodological, technical, and instrumental levels, FA has evolved substantially since the 1940s. Furthermore, these advances have finally also reached EFA, which is no longer a "second-class" option within the SEM family. If the type of problem and the nature of the data require EFA to be used, it can be used with the same standards of quality and rigor as any structural model.

Given this scenario, it would have been easy for me to criticize Yela's work from a purely technical, statistical, or "modern" view. After all, the maestro did not stop explaining the centroid method and graphical rotations from the beginning (Yela, 1956) to the end (Yela, 1997) of his career. And in his own research he tried advanced estimation procedures but usually ended up reporting principal axes estimates (e.g. Yela et al., 1969; García-Alcañiz & Yela, 1980; García-Jiménez & Yela, 1995; Muñiz & Yela, 1982). However, I would like to think I know better. Since statisticians took the lead, the advances in FA have become complex and technically demanding, but, at the same time, they have been implemented in programs that are more and more user-friendly and easy to use (well, some of them at least). And this contrast is conducive to creating a false sense of control and knowledge. My impression as a reviewer is that, in most cases, practitioners do not base the decisions they take on solid knowledge. Perhaps it would be a good idea to get back to basics and learn the fundamentals.

As elementary and repetitive as it may sound, the choice of methodological tools must be determined by the psychological problem one is trying to solve and by the nature of the data one is analysing. Other considerations are largely irrelevant. However, my impression is that in most cases the choices are based on the modernity or sophistication of the method, on the software available, or on pressure from ill-informed reviewers. As for this last remark, however, although it is easy to criticize reviewers, we should also consider that most of us play a double role as researchers and reviewers (he who is free from sin…).

The relevance of what I have said above is most important at present given the vast array of choices available. An FA can go from an independent-cluster, fully restricted solution, through an intermediate solution in which only a few key variables are specified as markers, to a totally unrestricted solution. Furthermore, in any of the chosen options, residuals can be allowed to correlate. In this panoply of choices, the researcher must consider that the more restrictions that are freed, the more flexible the model will become, and the more likely it is that statistical fit will be good, but at the same time, it will become more complex, unstable and difficult to interpret and replicate. So, the choices have to be determined a priori and well founded. Blind post-hoc modifications are likely to capitalize on change and help to discredit the technique and the researcher. Browne (2001) is quite clear in this respect: an EFA solution is always superior to a heavily post-hoc modified CFA solution (see also Gerbing & Hamilton, 1996).

My final comments will again be based on one of Yela's best known schemas. Generations of psychology students in Spain learned FA following the four basic phases in Yela's books: Preparation, Factoring, Rotating, and Interpretation (Yela, 1996b). And this schema is useful for comparing the situation of FA at the beginning of the period reviewed and in the present day. The spirit of the 1940s was optimistic: There were important psychological problems to be solved, and it was thought that FA would contribute to finding solutions. So, most of the effort was devoted to the first and the fourth phases: carefully thought-out designs and thoughtful interpretations. Where things faltered was in the technique: EFA was neither up to the ambition of the design nor allowed for interpretations as rigorous as was desirable.

In contrast, and in my view, the present state of affairs reflects a certain disenchantment, fatigue and routine. It is a scenario in which the undoubtable technical progress masks a scarcity of ideas, poor designs and superficial interpretations. As an involved party, I can predict that technical developments in FA will continue to evolve. Exactly how applications will evolve, however, I do not know, but I can summarize how I would like this evolution to be. It is time to re-pose relevant psychological problems, and to devote much of the research effort to planning good designs. The problem and the design would then determine the most appropriate technique to be selected (in Muñiz's words: as demanding and sophisticated as it takes), and the researchers themselves are responsible for acquiring proficiency in the required technique. One thing, however, is clear (at least to me); a good design analyzed with a modest technique that the researcher knows well is far superior to the latest and most sophisticated technique which the researcher does not fully understand, and which masks a poor, convenience design. As researchers, practitioners and reviewers, we should therefore reread Yela. This is the best recommendation I can make to close this article.

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 397-438. https://doi.org/10.1080/10705510903008204

Bock, R. D. (2007). Rethinking Thurstone. In Cudeck, R., & MacCallum, R. C. (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 35-46). Routledge.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*, 605-634. http://dx.doi.org/10.1146/annurev.psych.53.100901.135239

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*, 111-150. https://doi.org/10.1207/S15327906MBR3601_05

Butler, J. M. (1968). Descriptive factor analysis. *Multivariate Behavioral Research*, *3*, 355-370. http://dx.doi.org/10.1207/s15327906mbr0303_5

Cattell, R.B. (1986). The psychometric properties of tests: Consistency, validity and efficiency. In R.B. Cattell & R.C. Johnson (Eds.), *Functional Psychological Testing* (pp 54-78). Brunner/Mazel.

Cliff, N., & Hamburger, C. D. (1967). The study of sampling errors in factor analysis by means of artificial experiments. *Psychological Bulletin*, *68*, 430-445. https://doi.org/10.1037/h0025178

Cudeck, R., & MacCallum, R. C. (Eds.) (2007). *Factor analysis at 100: Historical developments and future directions*. Routledge.

Devlieger, I., & Rosseel, Y. (2017). Factor Score Path Analysis. *Methodology*, *13*, 31-38. https://doi.org/10.1027/1614-2241/a000130.

Ferrando, P. J., & Lorenzo-Seva, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: Some aspects of the problem and some suggestions. *Psicológica*, *21*, 301-323.

Ferrando, P. J., & Lorenzo-Seva, U. (2013). *Unrestricted item factor analysis and some relations with item response theory*. Department of Psychology, Universitat Rovira i Virgili. http://psico.fcep.urv.es/utilitats/factor

Ferrando, P. J., & Lorenzo-Seva, U. (2014). El análisis factorial exploratorio de los ítems: algunas consideraciones adicionales [Exploratory item factor analysis: Some additional considerations]. *Anales de Psicología, 30*, 1170-1175.

Ferrando, P. J., Lorenzo-Seva, U., & (2016). A note on improving EAP trait estimation in oblique factor-analytic and item response theory models. *Psicologica, 37*, 235-247.

Ferrando, P. J., & Lorenzo-Seva, U. (2017). Program FACTOR at 10: Origins, development and future directions. *Psicothema, 29*, 236-240. doi: 10.7334/psicothema2016.304

Ferrando, P. J., & Lorenzo-Seva, U. (2018). Assessing the Quality and Appropriateness of Factor Solutions and Factor Score Estimates in Exploratory Item Factor Analysis. *Educational and Psychological Measurement, 78*, 762-780. http://dx.doi.org/10.1177/0013164417719308

García-Alcañiz, E., & Yela, M. (1980). Dimensiones de la fluidez verbal en una muestra de adolescentes [Dimensions of verbal fluency in a teenager sample]. *Revista de Psicología General y Aplicada*, *35*, 127-138.

García-Jiménez, M. V., & Yela, M. (1995). Dimensiones factoriales de la fluidez verbal utilizando criterios cuantitativos y cualitativos [Factorial dimensions of verbal fluency using quantitative and qualitative criteria]. *Revista de Psicología General y Aplicada*, *48*, 571-580.

Gerbing, D. W., & Hamilton, J. G. (1996). Viability of exploratory factor analysis as a precursor to confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *3*, 62-72. https://doi.org/10.1080/10705519609540030

Guilford, J. P. (1952). When not to factor analyze. *Psychological Bulletin, 49*, 26-37. https://doi.org/10.1037/h0054935

Gustafsson, J. E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28*, 407-434. https://doi.org/10.1207/s15327906mbr2804_2

Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *The British Journal of Statistical Psychology, 8,* 65-81. https://doi.org/10.1111/j.2044-8317.1955.tb00321.x

Izquierdo, I., Olea, J., & Abad, F. J. (2014). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema, 26*, 395-400. https://doi.org/10.7334/psicothema2013.349

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32*, 443-482. http://dx.doi.org/10.1007/BF02289658

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183-202. http://dx.doi.org/10.1002/j.2333-8504.1967.tb00991.x

Jöreskog, K. G. (2007). Factor Analysis and its extensions. In R. Cudeck, & R. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 47-78). Routledge.

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh, 60*, 64-82. https://doi.org/10.1017/S037016460002006X

Lawley, D. N. (1944). The Factorial Analysis of Multiple Item Tests. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics, 62*, 74-82. https://doi.org/10.1017/S0080454100006440

Lawley, D. N., & Maxwell, A. E. (1963). *Factor analysis as statistical method*. Butterworth.

Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). Exploratory item factor analysis: A practical guide revised and updated. *Anales de Psicología*, *30*, 1151-1169. http://dx.doi.org/10.6018/analesps.30.3.199361

Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research, 34*, 347-356. https://doi.org/10.1207/S15327906MBR3403_3

Lorenzo-Seva, U. (2003). A factor simplicity index. *Psychometrika, 68*, 49-60. http://dx.doi.org/10.1007/BF02296652

Lorenzo-Seva, U., & Ferrando, P. J. (2020). Unrestricted factor analysis of multidimensional test items based on an objectively refined target matrix. *Behavior Research Methods, 52*, 116-130. https://doi.org/10.3758/s13428-019-01209-1

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85-110. http://dx.doi.org/10.1146/annurev.psych.53.100901.135239

McCrae, R. R., Zonderman, A. B., Costa, P. T., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology, 70*, 552-566. https://doi.org/10.1037/0022-3514.70.3.552

Mislevy, R.J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics, 11*, 3-31. http://dx.doi.org/10.2307/1164846

Mulaik, S. A. (1986). Factor analysis and Psychometrika: Major developments. *Psychometrika, 51*, 23-33. http://dx.doi.org/10.1007/BF02293996

Mulaik, S. A. (2010). *Foundations of factor analysis*. Chapman & Hall.

Muñiz, J. (1995). Reflexiones sobre la obra metodológica de Mariano Yela [Reflections on the methodological work of Mariano Yela]. *Revista de Psicología General y Aplicada*, *48*, 447-510.

Muñiz, J. (2021). Semblanza de Mariano Yela [Portrait of Mariano Yela]. *Psicothema, 33*(3), 353-355.

Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test [Ten steps for test development]. *Psicothema, 31*, 7-16. https://doi.org/10.7334/psicothema2018.291

Muñiz, J., & Yela, M. (1982). Modulación de la estructura factorial de la fluidez oral por variables cognoscitivas y de personalidad [Cognitive and personality variables as modulators of the factorial structure of oral fluency]. *Revista de Psicología General y Aplicada, 37*, 957-972.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115-132. https://doi.org/10.1007/BF02294210

Pascual, M., Ponsoda, V., Castellanos, R., & Seisdedos, N. (1997). Mariano Yela: aportaciones metodológicas en Psicología [Methodological contributions of Mariano Yela in Psychology]. *Estudios de Psicología, 18*, 61-69.

Rodríguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*, 223-237. https://doi.org/10.1080/00223891.2015.1089249

Siguan, M. (1994). En recuerdo de Mariano Yela (1921-1994) [Remembering Mariano Yela (1921-1994)]. *Anuario de Psicología*, *63,* 245-250.

Thurstone, L. L. (1947). *Multiple-factor analysis; a development and expansion of The Vectors of Mind*. University of Chicago Press.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1-10. http://dx.doi.org/10.1007/BF02291170

Tucker, L. R. (1983). Searching for structure in binary data. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement: A Festschrift for Frederic M. Lord* (pp. 215-235). Erlbaum.

Wolfle, D. (1940). Factor analysis to 1940. *Psychometric Monographs, 3,* 69.

Yela, M. (1949). Application of the concept of simple structure to Alexander's data. *Psychometrika*, *14*, 121-135. http://dx.doi.org/10.1007/BF02289148

Yela, M. (1956). *Psicología de las aptitudes: el análisis factorial y las funciones del alma* [Aptitude Psychology: Factor analysis and the functions of the soul]. Gredos.

Yela, M. (1968). La significación estadística de la estructura simple en el análisis factorial [Statistical significance of simple structure in factor analysis]. *Revista de Psicología General y Aplicada*, *23*, 313-323.

Yela, M. (1987). *Estudios sobre inteligencia y lenguaje* [Studies on intelligence and language]. Ediciones Pirámide.

Yela, M. (1996a). La forja de una vocación [The forging of a vocation]. *Psicothema*, *8*(Sup), 43-51.

Yela, M. (1996b). Los tests y el análisis factorial [Tests and factor analysis]. *Psicothema, 8*(Sup), 73-88.

Yela, M. (1997). *La técnica del análisis factorial* [The technique of factor analysis]. Biblioteca Nueva.

Yela, M., & García-Alcañiz, E. (1975). Fluidez verbal y personalidad [Verbal fluency and personality]. *Revista de Psicología General y Aplicada*, *30*, 1047-1058.

Yela, M., Pascual, M., & Diez, E. (1969). Dimensiones de la comprensión verbal [Dimensions of verbal comprensión]. *Revista de Psicología General y Aplicada, 99,* 626-627.