European Journal of Education Studies
ISSN: 2501 - 1111
ISSN-L: 2501 - 1111
Available on-line at: www.oapub.org/edu

doi: 10.5281/zenodo.3551953
Volume 6 | Issue 8 | 2019

TIME SERIES IN EDUCATIONAL PSYCHOLOGY: APPLICATION IN THE STUDY OF COGNITIVE ACHIEVEMENT

Cristiano Mauro Assis Gomes[1i],
Felipe Valentini[2]
[1]Universidade Federal de Minas Gerais,
Belo Horizonte, Brasil
[2]Universidade São Francisco,
Brasil

Abstract:
The study proposes the utility of time series methodology in the study of processes in educational psychology. Conjointly, the study applies time series in the study of cognitive achievement. Thirteen students from seventh to ninth grades performed an inductive reasoning test. The performance was measured in 20 different occasions and the observations were aggregated to generate a times series of 260 observations (20 different occasions by each individual multiplied by 13 participants). Result shows that a seasonal ARIMA (0,0,1) (1,1,1) adequately fits the data through a model comparative approach. Concluding, despites the complexity, ARIMA methodology is capable to investigate process, reducing the object of the study without lost its fundamental properties and dynamical aspects.

Keywords: time series; cognitive process; achievement measures; quantitative methods

1. Introduction

Understand the dynamic of processes is a valuable task in many fields, i.e. educational psychology (van Geert & Steenbeek, 2005a; van Geert & Fischer, 2009). However, the study of process tends to be very complex, demanding elegant strategies (Gomes, Ferreira & Golino, 2014; Gomes & Golino, 2015; Gomes, Araujo, Nascimento & Jelihovisch, 2018). One of those is time series quantitative methodology, that possesses a broad body of statistical techniques that can diminish the complexity of the object studied without losing the fundamental properties of the dynamic of processes (Van Geert & Steenbeek, 2008). As declared: "*The American Psychological Association's (APA's) Division 12 Task Force on Promotion and Dissemination of Psychological Procedures has explicitly recognized time-series designs as important methodological approaches that can fairly test treatment efficacy and/or effective psychological practice*." (Borckardt et al., 2008, p. 77).

[i] Correspondence: email cristianomaurogomes@gmail.com

Times series also can be helpful to model a test structure behind a dynamical process. Otherwise, a structure yielded by a study in a unique time might not be the best model throughout the entirely developmental process. Then, time series might be applied to access a structure in small samples or even in a unique participant, if the constructs were measured repeatedly.

Time series approach basically involves quantitative techniques that deal with data in a timepoint which correlated with the previous timepoints data. In other words, a datum in some occasion depends basically of the previous datum in time. This time dependence is named as autocorrelation. According to Van Geert and Steenbeek (2005b), this kind of dependency is the fundamental characteristic of any processual aspect, and time series, as other statistical tools, incorporates this postulate, which can be represented by the equation 1: $y_t = f(y_{t-1})$.

The value of y_t on a specific moment of time (t+1) is a function of its value on the previous moment (t). What equation 1 argues mathematically is that the state of any object studied is caused by the previous state of that object. Equation 1 contains the fundamental idea that it is possible to reduce the scope of an object without losing its processual characteristics (Van Geert & Steenbeek, 2005b). Time series methodology can do it because the time series axiom of temporal dependence is congruent with equation 1. In other words, the previous observations directly influence or cause the next observations. Despite the relevance of dependency in studies about process, traditional statistics approaches do not deal with this condition. An assumption of traditional statistics is the independence of the error in the data. The large majority of studies about the predictors of cognitive achievement apply those statistics, i. e intelligence (Gomes & Borges, 2007; Gomes & Borges, 2008; Gomes & Borges, 2009a, 2009b, 2009c; Gomes, 2010a, 2010b; Gomes, 2011b; Gomes, 2012; Gomes & Golino, 2012b; Golino & Gomes, 2012; Golino & Gomes, 2014; Muniz, Pasian, & Gomes, 2016), metacognition (Gomes, Golino & Menezes, 2014; Gomes & Golino, 2014; Pires & Gomes, 2017; Pires & Gomes, 2018), students' learning approaches (Gomes, 2010b; Gomes, 2011a; Gomes, Golino, Pinheiro, Miranda, & Soares, 2011; Gomes & Golino, 2012a; Gomes, 2013). By contrast, time series methodology is a class of techniques that considers and estimates the error dependency. Because of that, time series methodology brings new possibilities, but demands a considerable number of observations to estimate the parameters accurately, around 50 or more occasions of measurement (Glass, Willson & Gottman, 1975).

Velicer and Fava (2003) published an important paper that describes examples of time series in psychological research. Despite time series being more developed in areas such as engineering and economics, the creation of a class of models known as Autoregressive Integrated Moving Average (ARIMA) has aggregated the time series quantitative methodology to social sciences.

Because of the dependency, ARIMA time series estimates an autocorrelation matrix, which is the correlation between data in a moment of time and data about the same variable in previous moments of time (Coghlan, 2011). The autocorrelation matrix is accounted in terms of lags. In order to explain more concretely the autocorrelation and its relationship with lags, Table 1 shows a representation of how the data is organized to calculate lagged autocorrelation. Supposing that a person performed a IQ test 100 times, Table 1 shows the score of that person in the first 5 occasions. In the first time, the score was 100 points. The second score was 102 points, and so forth. To calculate the lagged one autocorrelation the data is organized in two columns. The left column, labeled IQ in Table 1, contains the data in the usual sequence in time. The right column, labeled IQ-1 in Table 1, contains the data corresponding to one moment before the usual sequence in time. So, the first score (100 points) in column IQ does not have any value to be matched with the column IQ-1. The second score (102 points) in column IQ is matched in column IQ-1 with the previous value (100 points) of column IQ. Therefore, in practical terms, lag one means the value that comes before the current value. Continuing, the third value in IQ column (103 points) is matched in IQ-1 column with the second value from the IQ column. That procedure goes way down to the last value of the IQ column is matched. The data arrangement of the lag two autocorrelation follows the same logic of the lag one autocorrelation. However, the data that will be matched with the usual sequence corresponds to two values before it. As can be observed in Table 1, considering the two columns: IQ and IQ-2: the first two values in the IQ column are not matched with any values in the column IQ-2. The third value (103 points) in the IQ column is matched with the first value from the IQ column. The first value (100 points) is two lags before the third value (103 points). The fourth value (105 points) in the IQ column is matched to IQ-2 column with the second value in IQ column (102 points), and so on.

Table 1: Representation of the data organization to calculate lagged auto-correlation

Time	IQ	IQ -1	IQ-2	IQ-3	…
	100	---	---	---	…
2	102	100	---	---	…
3	103	102	100	---	…
4	105	103	102	100	…
5	105	105	103	102	…
…	…	…	…	…	…

The ARIMA approach is a time series model. It is composed by the autoregressive component (AR or p), the integrated component (I or d), and the moving averages (MA or q). ARIMA integrates the simple AR and MA time series models. These models, as well as the integrated ARIMA, can be useful to identify a pattern of trend or seasonality, or both, behind a series of data. The trend indicates the growth follows a pattern throughout the series (for instance, a linear increase, or a quadratic decrease). The seasonality assumes the series shows variations into specifics intervals (for instance, years, or cycles of difficulties in a cognitive test). AR and MA try to capture the autocorrelations through the autoregressed lags (like the equation 1) and the errors of the lags $y_t = f\big((\mathcal{E}_t) + (\mathcal{E}_{t-1})\big)$

The autoregressive and the moving average components assume the series is stationary, which means the autocorrelations are the same for the entirely series and the data does not show trend or seasonality. If the time series is non-stationary, which is the predominant case, the trend from the time series must be partial out in ARIMA model. The name of this procedure is differentiation.

Then, the first task is verified if the time series is stationary. If that condition happens, the time series must be non-differentiated (d=0). However, stationarity usually is not encountered in times series because it demands that the trend mean level and the standard deviation would be equal throughout the entire series. It is very common that trend mean level increases or decreases in pieces of the series, as occurs with the standard deviation that can be smaller or bigger in different pieces of the series. To deal with a non-stationary series due a trend, one might apply a lag difference procedure, which consists in replacing the current value at time t for the difference of the previous one ($y_t - y_{t-1}$); ARIMA model automatically performs the difference procedure (parameter d). If the first differentiation achieves a stationary time series, then the parameter d is modeled to be one (d=1). If not, a second differentiation will be done and if achieves stationarity then d parameter must be modeled to be two (d=2). If the second differentiation still generates a non-stationary time series, a new differentiation is created, and so on. It is important to note the d=2 does not mean a differentiation in lag two, however the difference is performed in lag one and, then, applied again in lag one (i.e. d=2 means a double-differencing, and not a differentiation in lag two). It is also relevant to mention the parameter d is set up to deal with trend. If the data shows seasonality as well, it might be necessary do remove it setting another parameter D (capital D) larger than 0. In summary, when parameters d and D are different from 0, the ARIMA will remove the trend and the seasonality in an integrated procedure. Then, the ARIMA evolves parameters for trend (lowercase p, d, q) and seasonality (uppercase P, D, Q).

The oncoming steps evolve choosing the best model of autoregressive and moving averages. ARIMA demands the determination of a model that considers these three parameters (p,d,q). One manner to define the ARIMA parameters includes verifying the fit of each model from a class of different models, choosing some criteria as AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion). This is the comparative model fit approach. Other way is the graphic analysis approach. It evolves the calculation of the partial autocorrelations of observed time series and the observation of the pattern in a graphic. The partial autocorrelations are measure of the autocorrelations after the lags have been partial out.

The next task of the graphic analysis approach evolves select the new time series stationary obtained by differentiation of maintain the original times series if it was stationary. Next, observes if the partial autocorrelations drop abruptly after a certain lag, this indicates that the autoregressive component is predominant, so the parameter q (moving averages) must be modeled to be zero lags (q=0) and the parameter p must be modeled considering the last lag number that has a statistically significant value ($p \leq .05$). For example, if the two first lags of 90 occasions in the partial autocorrelations are statistically significant, p must be modeled equal to 2 lags (p=2). However, if the partial autocorrelations drop slowly to zero, the moving averages component is predominant, so the parameter p (autoregressive component) must be modeled to be zero lag (p=0) and parameter q must be modeled considering the last lag number with statistically significant value ($p \leq .05$) as the case presented about the p parameter.

2. Application of ARIMA

This study proposes apply time series ARIMA in a testing design that incorporates four cycles or recurrences of difficulty. The aim is to illustrate the use of time series ARIMA in the study of psychological processes, applying this approach to inspect the process related to the respondents' performance in the context of a specific testing design.

Explaining the testing design, the participants took a test to measure inductive reasoning. The items have been arranged in four cycles. Each cycle is composed by five levels of increased difficulty. When a cycle finishes with the all its five levels, the next one begins with the easiest level being presented first and so on. This process repeats until the end of the fourth cycle. Figure 1 shows the testing design. The difficulty of the items enhances until a peak, then easier items are presented, and the cycle starts over again. In this sense, we expect seasonality due the design of the items difficulty. First, the participants answer three items in the level one of difficulty (1-3 on the x-axis); then they answer three items with difficulty in the level two (4-6 on the x-axis); the process continues until the participants answer the three more difficult items (in the level five). In this moment, the first cycle ends and the second cycle begins. The procedure continues until the end of the fourth cycle. So, considering the exposed properties of time series quantitative methodology, it is expected that ARIMA approach will be capable to investigate the complex nature of the processual aspects evolved in the specified testing design. In this sense, we expect strong autocorrelation between the achievement in a specific time and the achievement in a previous time, which would justify modelling the data with time-series.

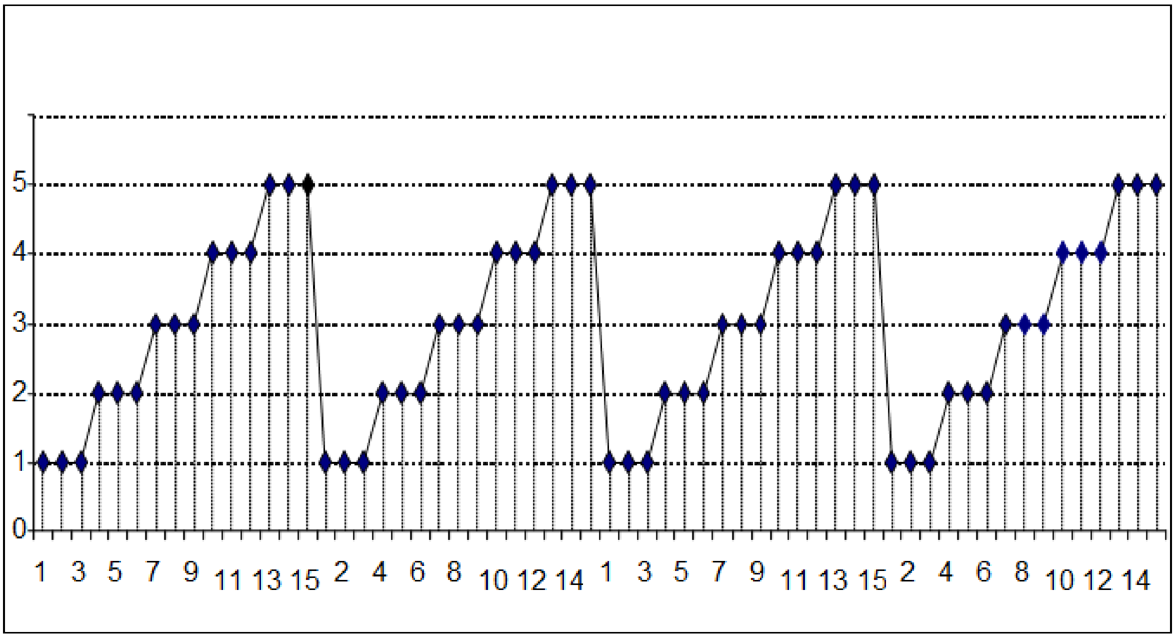

Figure 1: Testing Design: Difficulty Level (Y-axis) by Items (X-axis)

3. Method

3.1 Participants

Participants were 13 seventh-ninth grade students from a middle-class SES private school in the Belo Horizonte city, Minas Gerais, Brazil, seven male and six females, with mean age of 13.53 years (standard deviation of 1.05). Four of them were in seventh grade, three in eighth grade, and six in the ninth grade.

3.2 Instrument

3.2.1 Inductive Reasoning Developmental Test (IRDT)

IRDT (Golino, Gomes, Commons, & Miller, 2014) has 56 inductive reasoning items that contains seven difficult levels. The task included in all items is identifying from five patterns of letters the one that does not follow the same organization rule of the others (Golino & Gomes, 2012). Only the first five difficulty levels were employed in the current study.

Four booklet versions were created for the propose of this study. Each booklet version has 60 items, forming four cycles. Each cycle has five groups of three items, each one representing a specific level of difficulty. So, the first 15 items of the booklet compose the first cycle: items one, two and three measure the first difficulty level; items four, five, and six measure the second difficulty level, and so on, until items 13, 14 and 15 that measure the fifth difficulty level. The next 45 items compose the second, third and fourth cycle, organized with the same structure as cycle one. So, each booklet version is composed by four cycles (see Figure 1). Each cycle has one item that is also present in other cycle of the same booklet. The difference between the booklet versions is the order of the four cycles, e.g., the first cycle of the version one is the fourth cycle of the version four. The raw score is generated through the sum of right answers in each group of three items representing the difficulty levels in each cycle. Each group has a score that goes from 0 (participant failed all three items) to 3 (participant passed all three items). Then, each test taker had 20 raw scores (60 items / 3 difficulty items) gathered in four cycles, and each cycle has five levels of difficulty (then, 4 cycles X 5 difficulty = 20 raw scores). It is important to note Golino and Gomes (2012) show evidence the items on the same difficult level did have difficult invariant parameters (for instance, three first items of the first booklet have the same difficulty level of the three first items of the second booklet).

3.3 Procedures and Data Analysis

Each participant answered only one booklet version in a unique moment. Different participants responded different booklet versions, which variated due the order of the cycles. After the ending of one cycle of the booklet, it was given one minute of pause before the beginning of the next cycle.

The answers of each participant generate 20 raw scores of the achievement test. Each group of three items produces a raw score of zero, one, two or three points (respondent passed all three items). The 20 raw scores of the 13 participants were aggregated to compose a time series of 260 occasions. ARIMA time series was applied and the model was selected through the graphic analyzes of the autocorrelation matrix and partial autocorrelation matrix. Time series analysis was performed through the statistical program R version 3.02 (R Core Team, 2017).

4. Results and Discussion

Data was stored in a time series object through the *ts ( )* function from the R program (we switched the font to clearly distinguish a R function from the text). This part in syntax was *achievement <- ts (data, frequency=5)* what means that the data was transformed in a time series object in R program named achievement. The time series start with the first occasion and end with the 260 occasions. The frequency equal to five in the syntax represents the cycle of five different degrees of difficulty in the testing design. This means that each five measurement occasions compose a time cycle and that each lag in the model will represent a specific time cycle which is composed by five measurement occasions.

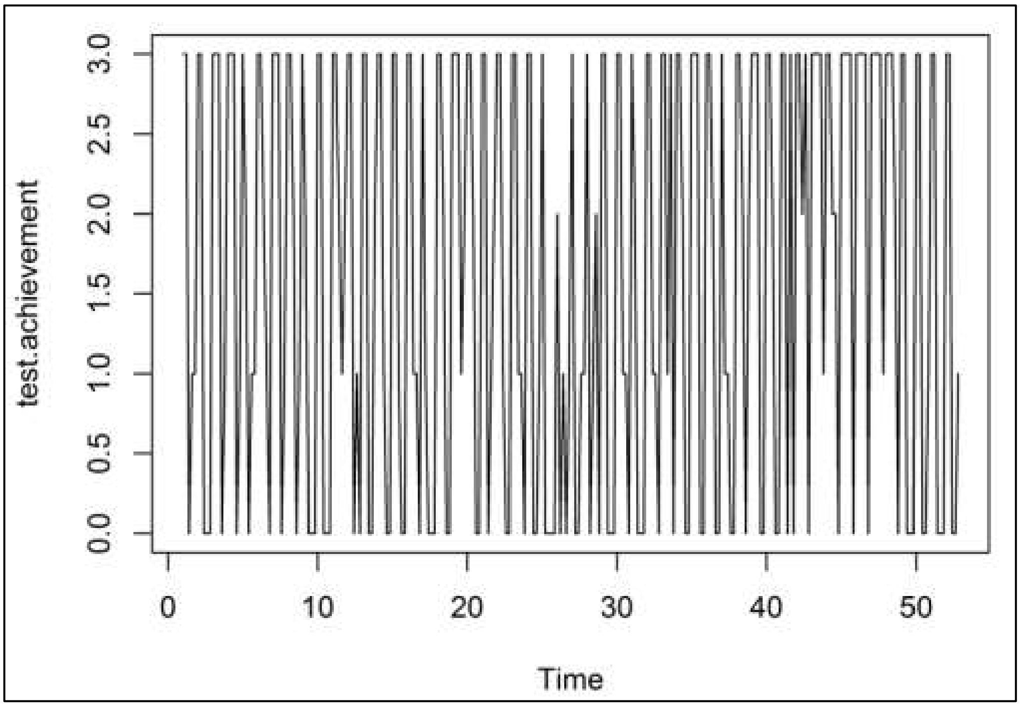

Figure 2: Observed time series for the achievement test

Figure 2 shows the observed time series plotted in a graphic. As commented, each time cycle represents five measurement occasions. And the cycles are compatible with a time-series with seasonality. Because time is accounted through the cycles, the 260 occasions are represented into 52 lag cycles because each cycle is composed by five occasions.

The mean level of the observed time series is 1.615, median is 2.000, standard deviation is 1.332, the minimum is 0 and maximum is 3. Standard deviation is large, indicating that the observed time series has a great variance. That condition is expected in function of the testing design of the study that presents to the participants easy item groups and hard item groups.

Figure 3 shows the autocorrelation factor of the observed time series. As mentioned, the time unit of the observed time series is the cycles. As each lag corresponds to each time unit, the first lag is composed by the lags 0.2, 0.4, 0.6, 0.8, and 1.0. Lag 0.2 is the lag between the first item group and the second item group into cycle one. Lag 0.4 is the lag between the first item group and the third item group into cycle one. Lag 0.6 is the lag between the first item group and the fourth item group into cycle one, and lag 0.8 is the lag between the first item group and the fifth item group into cycle one. Lag 1.0 represents the lag between the first item group from cycle one and the first item group from cycle two showing a correlation between two cycles of same difficult level. Continuing, the second lag is composed by the lags 1.2, 1.4, 1.6, 1.8 and 2.0. Lag 1.2 is the lag between the first item group into cycle one and the second item group into cycle two. Lag 1.4 is the lag between the first item group into cycle one and the third item group into cycle two. Lag 1.6 is the lag between the first item group into cycle one and the fourth item group into cycle two. Lag 1.8 is the lag between the first item group into cycle one and the fifth item group into cycle two. Lag 2.0 represents the lag between the first item group from cycle one and the first item group from cycle three. The same logic is so forth.

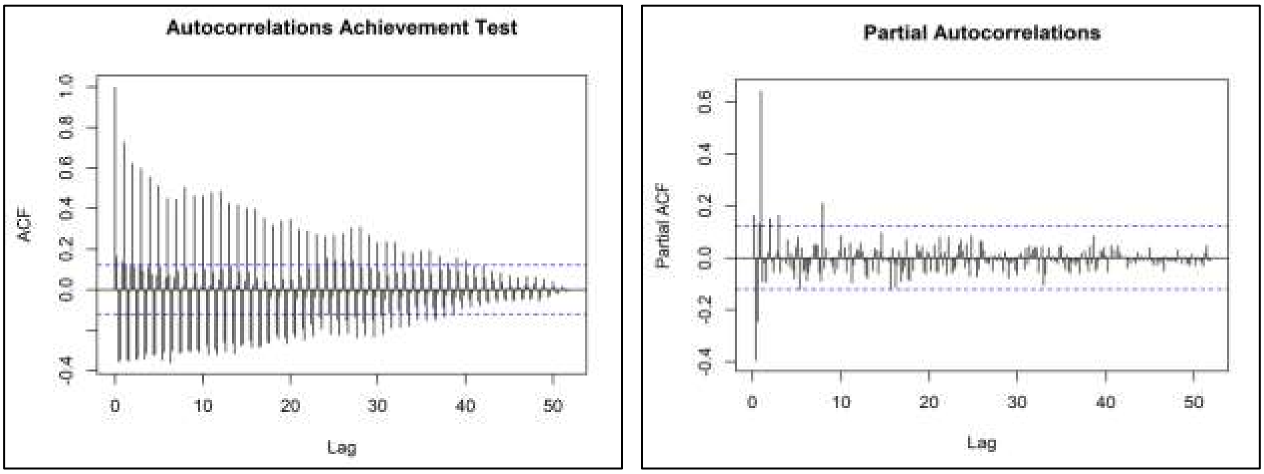

Figure 3: Autocorrelations and partial autocorrelations of the achievement test scores through lags

Working with ARIMA demands the identification of the model to be fitted. The first task is verifying if the time series is stationary. A strategy to do that is observe the autocorrelations pattern. One condition to verify if observed time series is not stationary is observe the autocorrelation lags. When the autocorrelations decay close to zero and after they after turn to rise, the time series is non-stationary. This is what happens with the autocorrelations of the observed time series of the study (Figure 3). The autocorrelations slow down close to zero and rises up, continually. An example of that pattern is encountered in the following cycle lags. The lag 0.2 has the autocorrelation of

.164. This indicates that the performance in the first item group is weakly correlated with the performance in the second item group, both from the first cycle of time series. However, the lag 0.4 has an autocorrelation of -.358 which indicates that the performance in the first item group correlates in a negative way with the performance in the third item group, both from the first cycle. The lag 0.6 has an autocorrelation of -.348, indicating that the performance in the first item group correlates negatively with the performance in the fourth item group, both from the first cycle. The autocorrelation from the lag 0.8 is .187 indicating that the performance in the first item group is weakly and positively correlated with the fifth item group, both from the first cycle. The lag 1.0 has an autocorrelation of .728 and presents a strong connection between the performance in the first item group from the first cycle with the performance in the first item group from the second cycle. The variety of positive and negative autocorrelations showed in lag 0.2, 0.4, 0.6, 0.8 and 1.0 is a recurrent pattern and indicates that the observed time series is non-stationary. Figure 3 shows that pattern in a graphic representation. The pattern of autocorrelations decays throughout the lags, as well as it shows sinusoid shape with peaks and valleys. It also means the series has likely seasonality, and an ARIMA model, with P, D, Q parameters, could fit to the data (Montgomery, Jennings & Kulahci, 2015).

There is a strong seasonal pattern. In this case, a seasonal ARIMA is applied in the time series, which has the three usual parameters p, d, q, and has three more parameters P, D, Q that represents the seasonal parameters. In this model, the six parameters need to be specified. Despite the more complexity, seasonal ARIMA is a good choice when the seasonal or cycle pattern is strong, which occurs in this study. Because the time series shows a non-stationary pattern, the first task is applying differencing until the time series achieve a stationary pattern. The differentiation can be done in d parameter or in D parameter to achieve stationarity. Some models which only estimate d and D parameters were compared through AIC criteria. Model 1 defines d=1 and D=0 (AIC=1017.86), model 2 defines d=0 and D=1 (AIC=702.65), model 3 states d=1 and D=1 (AIC=842.58), model 4 states d=2 and D=0 (AIC=1238.35), model 5 states d=0 and D=2 (AIC=938.13), model 6 states d=2 and D=2 (AIC=1329.72), model 7 defines d=1 and D=2 (AIC=1072.59) and finally model 8 defines d=2 and D=1 (AIC=1111.72). The model 2 is the best model because presents the lowest AIC value. This model was the chosen.

After the choice about the parameters d and D the rest of the parameters was analyzed. Parameters p, q, P, and Q were selected through the comparative fit approach. Table 2 shows the variety of models and the AIC value of each one. The seasonal ARIMA (0,0,1) (1,1,1) was the best model because presented the lowest AIC value and its parameters were all statistically significant ($p \leq .05$). The model applied differencing through the D parameter, which indicates seasonality, and it also means the time-series was non-stationary. Furthermore, the estimated parameters ma1, SMA1 (Seasonal Moving Average) and SAR1 (Seasonal Autoregressive) showed values ranging from -1 to 1 that denotes the model yielded stationarity properties. If those parameters range away from -1 to 1 the model would be non-stationary, and it would unsuitable for an ARIMA model.

Table 2: Fit of seasonal ARIMA models through AIC and its parameters

models (p,d,q) (P,D,Q)	**AIC**	**Parameters (and Standard Error)**							
		ar1	**ar2**	**ma1**	**ma2**	**SAR1**	**SAR2**	**SMA1**	**SMA2**
model 1 (1,0,0) (0,1,0)	700.67	0.125 (0.06)							
model 2 (2,0,0) (0,1,0)	702.46	0.128 (0.06)	-0.029 (0.06)						
model 3 (0,0,1) (0,1,0)	700.44			0.132 (0.06)					
model 4 (1,0,1) (0,1,0)	702.31	-0.141 (0.33)		0.270 (0.33)					
model 5 (0,0,2) (0,1,0)	702.32			0.130 (0.06)	-0.020 (0.06)				
model 1a (1,0,0) (0,1,1)	647.37	0.129 (0.06)						-0.616 (0.08)	
model 1b (1,0,0) (1,1,0)	672.93	0.118 (0.06)				-0.331 (0.06)			
model 1c (1,0,0) (1,1,1)	628.04	0.149 (0.06)				0.389 (0.07)		-0.974 (0.05)	
model 1d (1,0,0) (1,1,2)	629.13	0.142 (0.06)				0.562 (0.15)		-1.191 (0.19)	0.191 (0.18)
model 1e (1,0,0) (2,1,1)	629.52	0.145 (0.06)				0.382 (0.06)	0.049 (0.06)	-1.000 (0.18)	
model 3a (0,0,1) (0,1,1)	647.39			0.129 (0.06)				-0.615 (0.08)	
model 3b (0,0,1) (1,1,0)	672.91			0.120 (0.06)		-0.330 (0.06)			
model 3c (0,0,1) (1,1,1)	627.85			0.154 (0.06)		0.390 (0.07)		-0.974 (0.06)	
model 3d (0,0,1) (1,1,2)	629.02			0.146 (0.06)		0.557 (0.15)		-1.185 (0.19)	0.185 (0.18)
model 3e (0,0,1) (2,1,1)	629.38			0.149 (0.06)		0.383 (0.06)	0.047 (0.06)	-1.000 (0.18)	

The model 3c (Table 2) shows SAR and SMA (or P and Q) as the most important parameters, as they presented the greatest regression weights, which highlight the relevance of the seasonality properties to explain the data. The positive SAR1 indicates a positive correlation throughout cycles (or the previous cycle is positive associated to the current cycle); and the negative SMA1, the most weighted parameter in this model, points to a strong negative correlation between the end of a cycle and the beginning of the subsequently cycle. This result was expected due the strong difference in the difficulty between a group of items at the cycles edges. The ma1 shows less weight than the seasonal parameters (SAR and SMA), even though it can be interpreted as a slightly correlation from a group of items to the subsequent group, regardless the cycles. In summary, the ARIMA model could capture the theoretical process behind the instrument level of difficulty.

However, the AIC values presented in the Table 2 does not sustain, per se, that the model has an adequate fit but only says that the specific model is the best of the tested models. The model fit must be evaluated through the autocorrelation function of the residuals, the *p* values for Ljung-Box statistic and the normal distribution of the residuals. Figure 4 shows that there is no one significative autocorrelation of residual lags as there is no one *p* values for Ljung-Box statistic below .05, indicating that the null hypotheses that p is equal zero to the autocorrelation lags cannot be refuted. Beyond these aspects, the residuals show a normal distribution curve, indicating that the selected season ARIMA model represents adequately the observed time series.

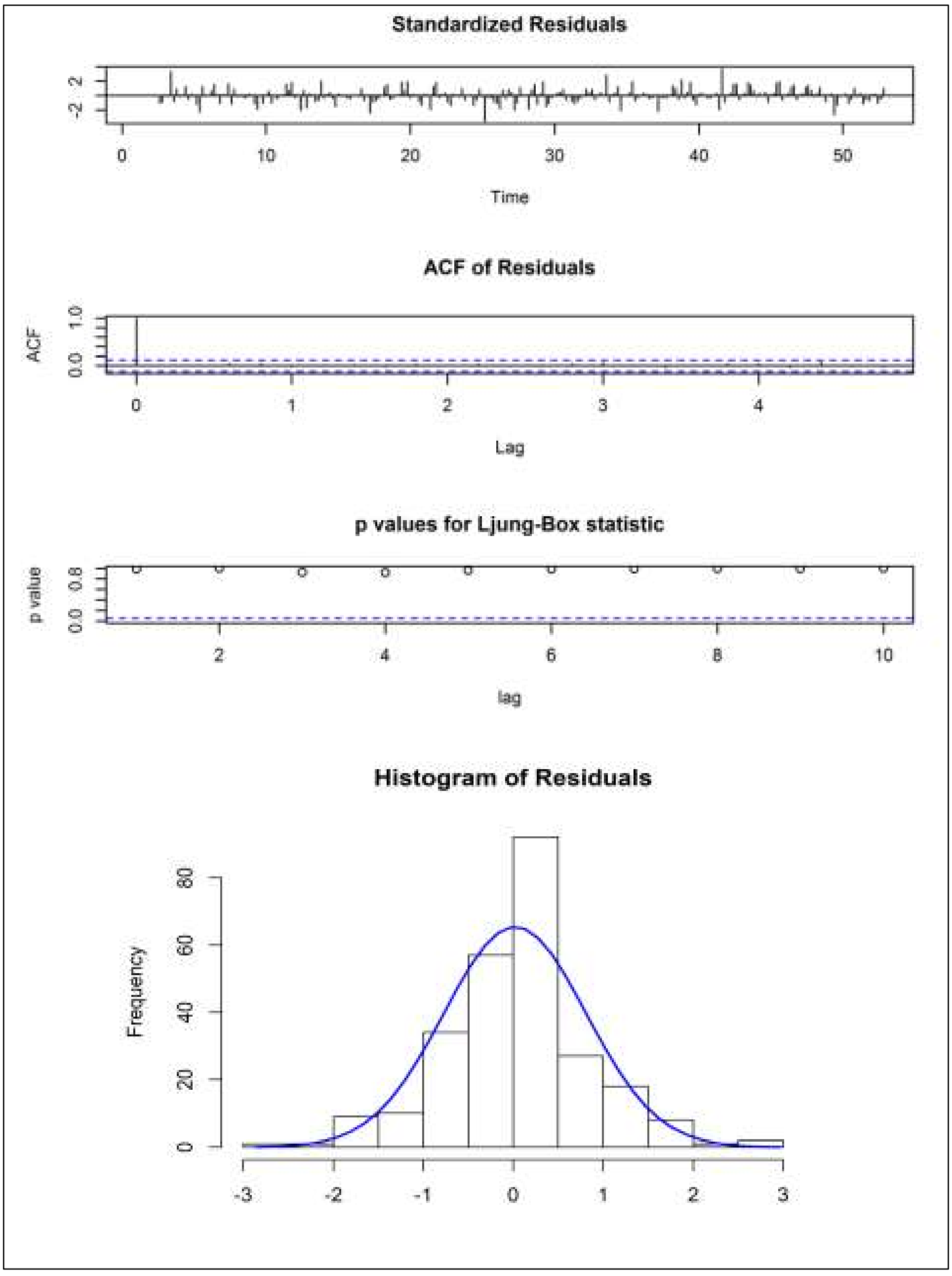

Figure 4: Fit tests of the selected seasonal ARIMA model

5. Conclusion

The present study proposes that ARIMA time series methodology is an elegant approach to deal with process in educational psychology. Concomitantly, the study applied a specific test design which each cycle was composed by five different item groups in terms of difficulty. The seasonal ARIMA was approached to the observed time series and adequate fit the data. This model showed the achievement in a time was slightly associated to the forthcoming achievement; the group of items at same level of difficult in different cycles are positive associated each other (for instance, first group of items of the first cycle was associated to the first group of item of the second cycle and so forth); and the strongest parameter pointed to an abrupt achievement decrease (or a valley on the time-series graphic) from the end of a cycle and the beginning of the subsequent cycle. Those results yield relevant information about the achievement process and the testing designing due the difficulty of the items. Those results also support the hypotheses of a time-series with seasonality. In this sense, the model is an evidence of temporal dependence of the achievement test (i.e. an achievement score depends on the previous score). However, it does not highlight any cognitive process behind the achievement.

Despite the complexity of investigate process, time series quantitative methodology is a viable tool and does not eliminates the fundamentals and the dynamics of the processes. Of course, this study applies ARIMA methodology to only one variable and other aspects could be part of the research, as exogenous factors and other time series in a complex fashion evolving multivariate time series analysis. Beyond the proximal objectives, this study intended stimulate the increase of psychological research with time series in fields that traditionally do not use this approach, as the intelligence psychometric field, for example. Through time series quantitative approach new possibilities can be open to these fields, articulating structures and processes together.

Acknowledgements
Cristiano Mauro Assis Gomes has a productivity scholarship awarded by CNPq, Brazil (level 2).

References

Borckardt, J. J., Nash, M. R., Murphy, M. D., Moore, M., Shaw, D., & O'Neil, P. (2008). Clinical practice as natural laboratory for psychotherapy research: a guide to case-based time-series analysis. *American Psychologist*, 63 (2), 77-95. doi: 10.1037/0003-066X.63.2.77

Coghlan, A. (2011). *A little book of R for time series.* Retrieved from https://github.com/avrilcoghlan/LittleBookofRTimeSeries/raw/master/_build/latex/TimeSeries.pdf.

Glass, G. V., Willson, V. L., & Gottman, J. M. (1975). *Design and analysis of time series experiments.* Boulder, CO: Colorado Associate University Press.

Golino, H. F. & Gomes, C. M. A. (2012). The structural validity of the Inductive Reasoning Developmental Test for the Measurement of Developmental Stages. In *Adult Development: Past, Present and New Agendas of Research*. Symposium conducted at the meeting of the European Society for Research on Adult Development, Coimbra, PT. Retrieved from http://www.europeadultdevelopment.org/materials1.html.

Golino, H. F., & Gomes, C. M. A. (2014). Psychology data from the "BAFACALO project: The Brazilian Intelligence Battery based on two state-of-the-art models – Carroll's Model and the CHC model". *Journal of Open Psychology Data*, *2*(1), p.e6. doi: 10.5334/jopd.af

Golino, H. F., Gomes C. M. A., Commons, M. L., & Miller, P. M. (2014). The construction and validation of a developmental test for stage identification: Two exploratory studies. *Behavioral Development Bulletin, 19* (3), 37-54.

Gomes, C. M. A. (2010a). Avaliando a avaliação escolar: notas escolares e inteligência fluida. *Psicologia em Estudo, 15*(4), 841-849. Recuperado de http://www.redalyc.org/articulo.oa?id=287123084020

Gomes, C. M. A. (2010b). Estrutura fatorial da Bateria de Fatores Cognitivos de Alta-Ordem (BaFaCalo). *Avaliação Psicológica, 9*(3), 449-459. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712010000300011&lng=pt.

Gomes, C. M. A. (2010c). Perfis de Estudantes e a relação entre abordagens de aprendizagem e rendimento Escolar. *Psico (PUCRS. Online), 41*(4), 503-509. Recuperado de http://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico/article/view/6336

Gomes, C. M. A. (2011a). Abordagem profunda e abordagem superficial à aprendizagem: diferentes perspectivas do rendimento escolar. *Psicologia: Reflexão e Crítica*, *24*(3), 438-447. doi: 10.1590/S0102-79722011000300004

Gomes, C. M. A. (2011b). Validade do conjunto de testes da habilidade de memória de curto-prazo (CTMC). *Estudos de Psicologia (Natal), 16*(3), 235-242. doi:10.1590/S1413-294X2011000300005

Gomes, C. M. A. (2012). Validade de construto do conjunto de testes de inteligência cristalizada (CTIC) da bateria de fatores cognitivos de alta-ordem (BaFaCAlO). *Gerais : Revista Interinstitucional de Psicologia*, *5*(2), 294-316. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1983-82202012000200009&lng=pt&tlng=pt.

Gomes, C. M. A. (2013). A Construção de uma Medida em Abordagens de Aprendizagem. *Psico (PUCRS. Online), 44*(2), 193-203. Recuperado de http://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico/article/view/11371

Gomes, C. M. A., & Borges, O. N. (2007). Validação do modelo de inteligência de Carroll em uma amostra brasileira. *Avaliação Psicológica*, *6*(2), 167-179. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712007000200007&lng=en&tlng=pt.

Gomes, C. M. A., & Borges, O. (2008). Qualidades psicométricas de um conjunto de 45 testes cognitivos. *Fractal: Revista de Psicologia*, *20*(1), 195-207. doi:10.1590/S1984-02922008000100019

Gomes, C. M. A. & Borges, O. N. (2009a). O ENEM é uma avaliação educacional construtivista? Um estudo de validade de construto. *Estudos em Avaliação Educacional*, *20*(42), 73-88. doi: 10.18222/eae204220092060

Gomes, C. M. A.s, & Borges, O. N. (2009b). Propriedades psicométricas do conjunto de testes da habilidade visuo espacial. *PsicoUSF*, *14*(1), 19-34. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-82712009000100004&lng=pt&tlng=pt.

Gomes, C. M. A., & Borges, O. (2009c). Qualidades psicométricas do conjunto de testes de inteligência fluida. *Avaliação Psicológica*, *8*(1), 17-32. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712009000100003&lng=pt&tlng=pt.

Gomes, C. M. A., & Borges, O. (2008). Limite da validade de um instrumento de avaliação docente. *Avaliação Psicológica*, *7*(3), 391-401. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712008000300011&lng=pt&tlng=pt.

Gomes, C. M. A., Golino, H. F., Pinheiro, C. A. R., Miranda, G. R., & Soares, J. M. T. (2011). Validação da Escala de Abordagens de Aprendizagem (EABAP) em uma amostra Brasileira. *Psicologia: Reflexão e Crítica*, *24*(1), 19-27. doi: 10.1590/S0102-79722011000100004

Gomes, C. M. A., & Golino, H. F. (2012a). Incremental validity of the Learning Approaches Scale. *Psicologia: Reflexão e Crítica*, *25*(4), 623-633. doi: 10.1590/S0102-79722012000400001

Gomes, C. M. A., & Golino, H. F. (2012b). O que a inteligência prediz: diferenças individuais ou diferenças no desenvolvimento acadêmico? *Psicologia: teoria e prática*, *14*(1), 126-139. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-36872012000100010&lng=pt&tlng=pt.

Gomes, C. M. A., & Golino, H. F. (2014). Self-reports on students' learning processes are academic metacognitive knowledge. *Psicologia: Reflexão e Crítica*, *27*(3), 472-480. doi: 10.1590/1678-7153.201427307

Gomes, C. M. A., Golino, H. F., & Menezes, I. G. (2014). Predicting School Achievement Rather than Intelligence: Does Metacognition Matter? *Psychology, 5*, 1095-1110. doi: 10.4236/psych.2014.59122

Gomes, C. M. A., de Araújo, J., Ferreira, M. G., & Golino, H. F. (2014). The validity of the Cattel-Horn-Carroll model on the intraindividual approach. *Behavioral Development Bulletin, 19*(4), 22-30. doi: 10.1037/h0101078

Gomes, C. M. A., & Golino, H. (2015). A retenção fatorial na abordagem intraindividual: proposição de uma estratégia de triangulação. *Avaliação Psicológica*, *14*(2), 273-

279. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712015000200013&lng=en&tlng=pt.

Gomes, C. M. A., Araujo, J., Nascimento, E., & Jelihovisch, E. (2018). Routine Psychological Testing of the Individual Is Not Valid. *Psychological Reports, 122*(4), 1576-1593. doi: 10.1177/0033294118785636

Montgomery, D. C., Jennings, C. L., Kulahci, M. (2015). Introduction to time series analysis and forecasting (2nd. ed.). Hoboken, NJ: Wiley.

Muniz, M., Gomes, C. M. A., & Pasian, S. R. (2016). Factor structure of Raven's Coloured Progressive Matrices. *Psico-USF*, *21*(2), 259-272. doi: 10.1590/1413-82712016210204

Pires, A. A. M., & Gomes, C. M. A. (2018). Proposing a method to create metacognitive school exams. *European Journal of Education Studies, 5*(8), 119-142. doi:10.5281/zenodo.2313538

Pires, A. A. M., & Gomes, C. M. A. (2017). Three mistaken procedures in the elaboration of school exams: explicitness and discussion. *PONTE International Scientific Researches Journal, 73*(3), 1-14. doi: 10.21506/j.ponte.2017.3.1

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL https://www.R-project.org/.

Van Geert, P. & Steenbeek, H. (2005a). The dynamics of scaffolding. *New Ideas in Psychology*, 23, 115-128. doi:10.1016/j.newideapsych.2006.05.003

Van Geert, P. & Steenbeek, H. (2005b). Explaining after by before. Basic aspects of a dynamic systems approach to the study of development. *Developmental Review*, 25 (34), 408-442. doi.org/10.1016/j.dr.2005.10.003

Van Geert, P., & Steenbeek, H. (2008). A complexity and dynamic systems approach to developmental assessment, modeling and research. In A. M. Battro, K. W. Fischer & P. Léna (Eds.). *The educated brain: essays in neuro-education* (pp. 71-94). Cambridge, UK: Cambridge University Press.

Van Geert, P. L. C., & Fischer, K. W. (2009). Toward a New Grand Theory of Development? Connectionism and Dynamic Systems Theory Reconsidered. In Spencer, J. P., Thomas, M. S. C. & McClelland, J. (Eds.). *Dynamic systems and the quest for individual-based models of change and development* (pp. 313-336). Oxford, UK: Oxford University Press.

Velicer, W. F., & Fava, J. L. (2003). Time Series Analysis. Em J. Schinka & W. F. Velicer (Eds.), Research Methods in Psychology (pp. 581-606). Volume 2, Handbook of Psychology. New York, NY: John Wiley & Sons.