How to read and review an empirical paper

Questions you can/should address in your reports:

Research question and hypothesis:

- Is the researcher focused on well-defined questions?

- Is the question interesting and important?

- Are the propositions falsifiable?

- Has the alternative hypothesis been clearly stated?

- Is the approach inductive, deductive, or an exercise in data mining? Is this the right structure?

Research design:

- Is the author attempting to identify a causal impact?

- Is the "cause" clear? Is there a cause/treatment/program/fist stage?

- Is the relevant counterfactual clearly defined? Is it compelling?

- Is the method for doing so clear and compelling? Has statistical inference been confused with causal inference?

- Does the research design identify a very narrow or a very general source of variation?

- Could the question be addressed with another approach?

- Useful trick: ask yourself, "What experiment would someone run to answer this question?"

Theory/Model:

- Is the theory/model clear, insightful, and appropriate?

- Could the theory benefit from being more explicit, developed, or formal?

- Are there clear predictions that can be falsified? Are these predictions "risky" enough? Does the theory generate any prohibitions that can be tested?

- Would an alternative theory/model be more appropriate?

- Could there be alternative models that produce similar predictions—that is, does evidence on the predictions necessarily weigh on the model or explanation?

- Is the theory a theory, or a list of predictions?

- Is the estimating equation clearly related to or derived from the model?

Data:

- Are the data clearly described?

- Is the choice of data well-suited to the question and test?

- Are there any worrying sources of measurement error or missing data? Are any proxies reasonable?

- Are there sample size or power issues?

- Could the data sources or collection method be biased?

- Are there better sources of data that you would recommend?

- Are there types of data that should have been reported, or would have been useful or essential in the empirical analysis?

Empirical analysis:

- Are the statistical techniques well suited to the problem at hand?

- What are the endogenous and exogenous variables?

- Has the paper adequately dealt with concerns about measurement error, simultaneity, omitted variables, selection, and other forms of bias and identification problems?

- Is there selection not just in who receives the "treatment", but in who we observe, or who we measure?

- Is the empirical strategy convincing?

- Could differencing, or the use of fixed effects, exacerbate any measurement error?

- Did the author make any assumptions for identification (e.g. of distributions, exogeneity, etc)?

- Were these assumptions tested and, if not, how would you test them?

- Are the results demonstrated to be robust to alternative assumptions?

- Does the disturbance term have an interpretation, or is it just tacked on?

- Are the observations i.i.d., and if not, have corrections to the standard errors been made?

- What additional tests of the empirical strategy would you suggest for robustness and confidence in the research strategy?

- Are there any dangers in the empirical strategy (e.g. sensitivity to identification assumptions)?

- Can you imagine a better, or alternative, empirical strategy?

Results:

- Do the results adequately answer the question at hand?

- Are the conclusions convincing? Are appropriate caveats mentioned?

- What variation in the data identifies the elements of the model?

- Are there alternative explanations for the results, and can we test for them?

- Could the author have taken the analysis further, to look for impact heterogeneity, for causal mechanisms, for effects on other variables, etc?

- Is absence of evidence confused with evidence of absence?

Scope:

- Can we generalize these results?

- Has the author specified the scope conditions?

- Have casual mechanisms been explored?

- Are there further types of analysis that would illuminate the external validity, or the causal mechanism at work?

- Are there other data or approaches that would complement the current one?